Formula One Pit-Stop Strategy Data Analysis

香 港 大 學

THE UNIVERSITY OF HONG KONG

COMP4801

Final Year Project

Name: Ming Chung Chan, Jack

UID: 3035373169

Name: Chan Ming Chung
UID: 3035373169

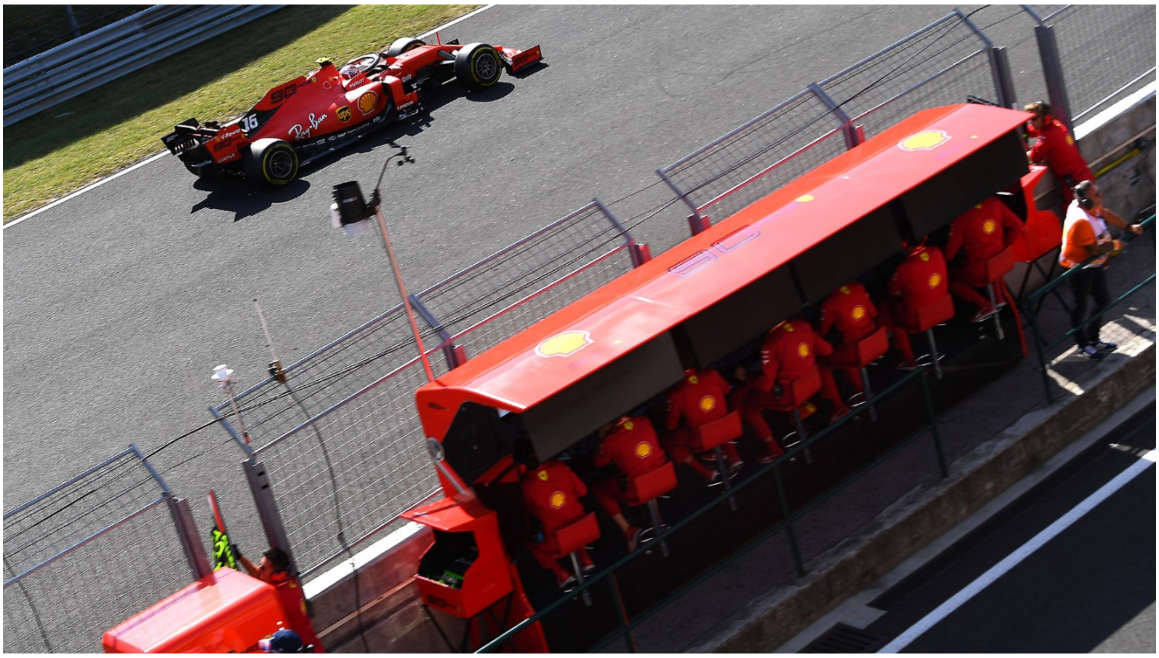
Table of Contents

Name: Chan Ming Chung
UID: 3035373169

Project Background

Formula One: The king of motorsport

Formula One has been the most prestigious motorsport racing event worldwide for 70 years. “Formula” refers to a set of technical regulations enforced for this specific type of single-seater racing cars governed by the FIA (or International Automobile Federation (Bekker & Lotz, 2009). As mechanical technologies advance with an increasing abundance of financial resource support, the rules and regulations of Formula One evolve to make the sport more competitive. While the constructors compete by developing their engine and chassis with the latest technologies, real-time race simulations are also conducted back in the factories to provide useful insights that could greatly affect team decision-making before and during the race. The results are sent to the pit wall, where the race engineers make pit stop strategy decisions and communicate with the drivers and all pit crew members.



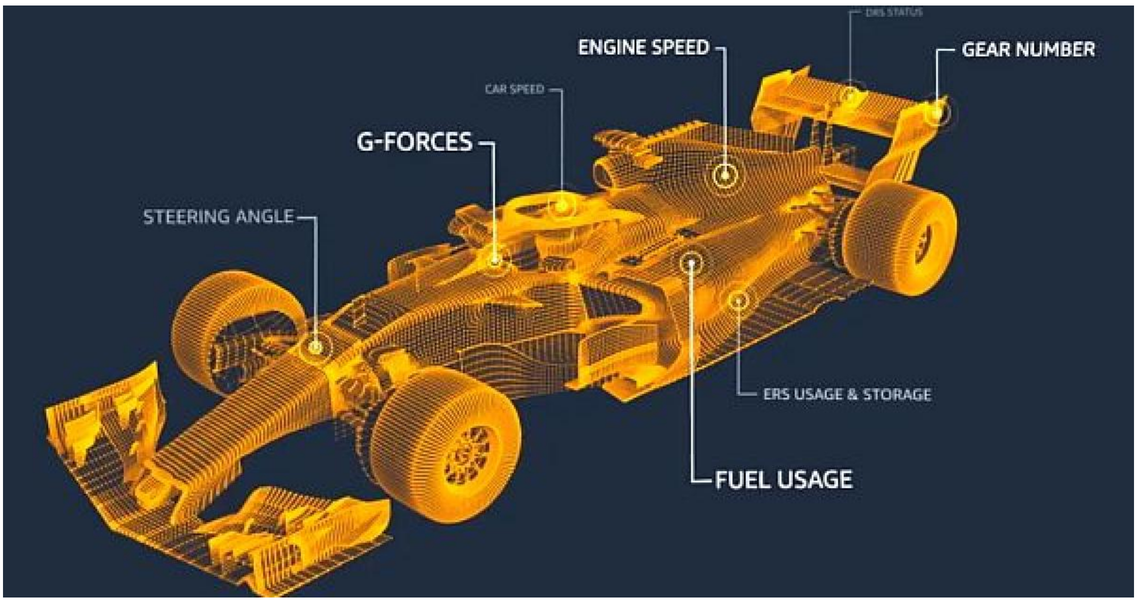
Scuderia Ferrari pit wall, with Team Principal Mattia Binotto
in the middle, and 6 race engineers/strategists.

The Decision to publicize data

Recently in 2019, the FIA (F1 official) has generously made racing data available, a move that immensely lowers the barrier to data mining and analytics from the public. Supported by motorsportstats.com, the online database gives numeric details of every historical race, such as lap time, pit stop decisions, driver and constructor championship scores, circuit and weather data, etc. Along with trustworthy third-party data providers online such as STATS F1, Ergast, Kaggle, and more, we could gather sufficient pieces of data that formulate a digital version of every F1 race.

Analytics application development

In this multi-million-dollar industry, a digital war has never been fiercer among back-end factories of each constructor team. Although it is widely known that every competitor possesses such facilities that serve as strategy suggesters, most of the details of the simulation machines and algorithms are hidden from publicity for two apparent reasons: fear of being copied by competitors, and the enormous amount of resources put into the research and development.

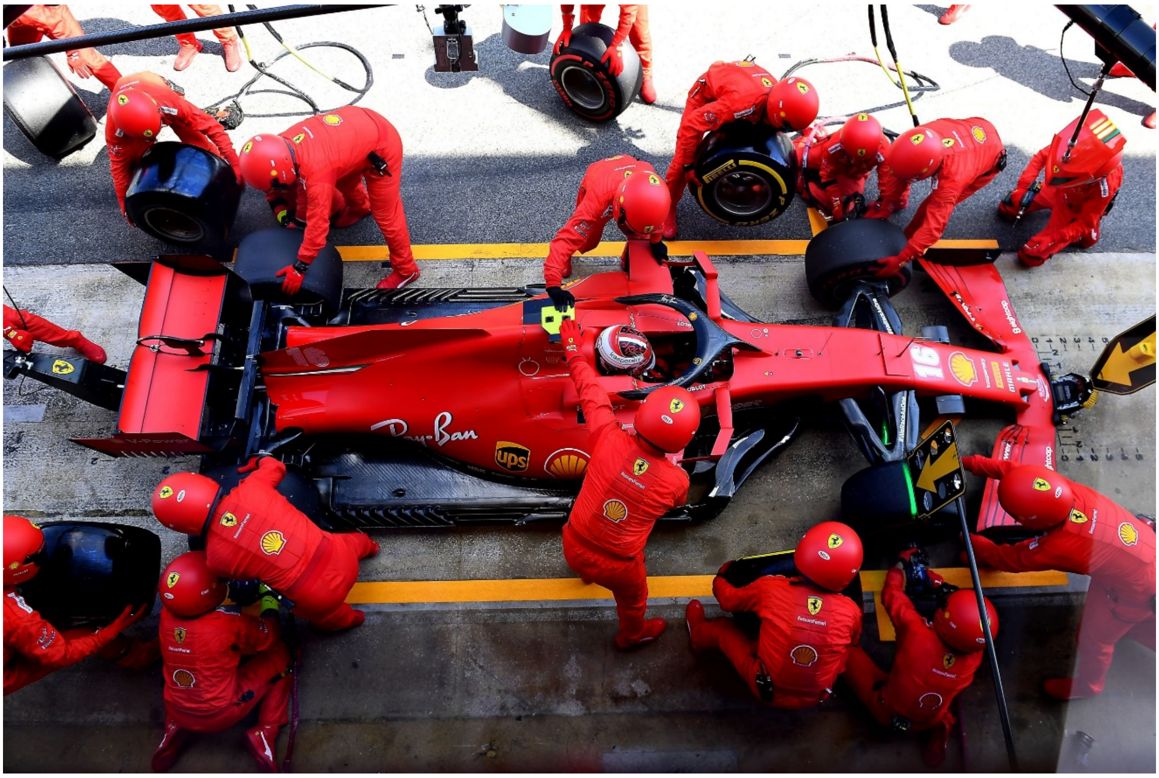


F1 Insights, powered by AWS.

In light of increasing audience engagement, F1 engaged with the AWS team to provide real-time analytical summaries graphically to let the audience gain effectively more insights about the race. Examples of insights include chasing distance between two competitors, predicted pit-stop strategy, and the probability of overtaking. This opens a new door to transform how audiences traditionally watch a car race, by feeling how close a car chase is even if the two cars are a pit stop apart (at approximately 30 seconds apart from each other). Increasing effort in new features and improvement in prediction accuracy has been seen in the 2019 and 2020 seasons as AWS takes bold strides on committing consistent developments.

Motivation, and Value of this project

In the past decade, top teams such as Mercedes Benz, Scuderia Ferrari, Red Bull, and McLaren F1 Racing have committed huge and fatal pit-stop strategy calls, costing them victories or even the championships. Infamous instances include the 2012 Abu Dhabi GP Alonso's too-early pit-stop, 2015 Monaco GP Hamilton too-late pit-stop, and the list goes on. When car competitiveness and driver skill level have marginal distinctness, the pit-stop strategy can be very decisive, if not determinant, to the race result. A good pit-stop allows overtaking without even a high-risk close car fight, while a bad pit-stop could put the car in traffic, harming the tyres and the possibility to make overtaking maneuvers.



Scuderia Ferrari Pit-stop for Charles Leclerc, Spanish Grand Prix 2020.

The major motivation of this project stems from the persistently unsatisfactory performance in the racing strategy of the Scuderia Ferrari F1 Team throughout recent seasons. While the analytical prediction of tyre degradation, gas consumption, and checking of car condition perform up to standard, Ferrari passively reacts to racing incidents and makes slow pit-stop decisions after observing how the other teams react. This project serves as a plug-and-use system for Scuderia Ferrari as a useful reference to swiftly form pitstop decisions by considering all race factors. The project should also standalone be a general prediction model of race result if it is fed with real-time racing data during a race.

Literature Review

In view of previous academic research on Formula One strategy, there are only a few initiatives in attempting to predict the final result of a race by providing live data. Planning Formula One race strategies by discrete-event simulation were proposed to recreate the race environment through a time-based approach (Bekker & Lotz, 2009). The model simulated a variety of in-race events, including car failures, passing maneuvers and pit stops, by using input data such as lap times, fuel load, position changes, etc. The model is run on Forza Motorsport 2, a video game in 2008, such that the model was contained in a human intervention-free environment. While the model demonstrated how the use of such publicly available data could create an accurate prediction to the 3 chosen test races, there were some disadvantages of such an approach, which would cause it less compatible with our project: Firstly, using an AI-controlled gaming environment for simulation may lose realism and neglect important factors from the real race. Next, fuel refilling was obsolete in the modern era, which existed in the 2005 season. Finally, only positive results from 3 test races were provided to conclude the model's high accuracy, which was not convincing enough.
A new approach in determining the optimal race strategy was proposed, by customizing several statistical models based on quadratic regression models, linear regression models, etc., that considered factors that could contribute to final gain in each race strategy (Sulsters & Bekker, 2018). It also included probability estimation on DNF (or Did-Not-Finish) probability for each racer using Bayesian inference. Such a statistical approach was tested against 4 races in the 2016 season and proved that the model predicted well in the end positions of racers, and the number of successful overtaking actions, but badly in the total race times. The model introduced a wide range of statistical models which take quantified

data for running the predictions, which could help generalize the race environment and justify the prediction results by varying the controllable input data, instead of running a gameplay scenario for each prediction at the previous model. However, the model came with several flaws: To start with, there was a strong assumption of a two-stop pit stop strategy must be used for every racer. Such an assumption ignored the one-stop strategy which gained increasing popularity in the modern F1 era and was proven to be more optimized for some tracks. Another issue stems from the casually assumed data which was understandably hard and impractical to collect, such as the fuel level data for each lap of a racer. The model tended to assume a linear relationship of data in a time-series.

The last literature shed light on how DSS (or Decision Support System) could both benefit and jeopardize the decision-making process of racing strategies. It gave a detailed analysis on the Abu Dhabi 2010 race, where Scuderia Ferrari driver Fernando Alonso unexpectedly lost the Formula One Driver World Championship due to the team's mistaken race strategy (Aversa, Cabantous & Haefliger, 2018). With extra abundant financial resources, Ferrari used a Monte Carlo simulation model with self-designed deterministic parameters, which was programmed to feedback two options regarding pit stop strategies. It was concluded with three major mistakes stemming from the DSS: Firstly, the model suffered from temporal distortion by the inability to update weight parameters by the fixed basic assumptions and underlying parameters. Secondly, Ferrari's strong chain of command and hierarchical management structure discouraged bottom-up interventions. Problems and situations from the real race event might not be reported and fully reflected in the simulation model. Lastly, the performativity of DSS was overlooked. Chris Dyer, the head of Strategy at Ferrari explained that the model gave option A and B, and he correctly chose A because the predicted result of A was optimal, but actually they miss out on option C which is not suggested from the DSS. Together with the fact that there was a clear rule stating that the output of DSS must not be disregarded, whenever the DSS made a prediction error the result could be catastrophic.

To conclude the literature review, we should focus on a model that generalizes quantified data and make predictions that is accurate in various race event environments. The model should also be treated as a reference with several optimized strategy suggestions instead of over-relying on which. We also have to treat data carefully and consider testing our model against a larger subset of race event data before evaluating the model’s accuracy. Finally, we have to pay attention to characteristics of the modern Formula One era, such as the race regulations and common pit stop strategy patterns.

Project Objective

Main Objective

In this project, the main objective is to build a system that is trained with historical race data to predict the optimal pit-stop strategy for the two drivers in the Scuderia Ferrari F1 Team. The system should consider a wide range of factors that are collected in real-time, so as to dynamically generate a prediction about the position when the race evolves, hence optimizing the race results of the 2 Ferrari drivers by selecting the best pit-stop strategy with the least time spent on any human decision-making processes.

Since we have to evaluate the accuracy of our model, our next goal is to use recent F1 seasons as testing data and reach an accuracy rate of >= 70% of predicting final track position after adopting a certain pit-stop strategy.

Project Methodology

Data Source and Data Collection

This project relies on several main data sources, including FIA F1 official website, STATS F1, Ergast, McLarenf-1, and also Kaggle for open-source data sources. F1 TV subscription provides telemetry data and traditional live broadcast services for live real-time data analytics. General relational database management software will be adopted for data storage after conducting data cleansing.

Machine Learning

This project adopts machine learning for precise predictions and constant refinement of the prediction model. Feeding the learning model with previous race data, it should create a balanced equilibrium of all factors influencing pit-stop strategy. Tensorflow Keras API is planned to facilitate this task. We will use the most recent seasons for the initial stage since they have the closest relevancy to the regulations of the current season, such as regulations on car design, fuel usage, track design, etc. The project will then extend to older seasons up until 2016 season towards the final phase.

Experiments and Results

Machine Learning Model 1: Predict final position gain or loss

We adopted Tensorflow Keras Sequential model for predicting position gained or lost at the end of the race when given a set of factors as input. This is a supervised regression type machine learning model where we feed the model with 17 input factors and 1 result label about the gain or loss in the final position.

Dataset description

The list of input factors are as follows:

1. Initial starting position of the racer.
2. Number of pit stop made in the race.
3. Starting tyre compound choice.
4. Tyre travel distance of starting tyre.
5. Second tyre compound choice.
6. Tyre travel distance of second tyre.
7. Third tyre compound choice.
8. Tyre travel distance of third tyre.
9. Number of pit stops under safety car or yellow flag conditions.
10. Track temperature.
11. Track humidity.
12. Track maximum altitude change.
13. Track number of turns in a lap.
14. Track race distance / total number of laps.

15. Track total length of race (in km).

16. Team Ability index (from previous season constructors' championship points).

17. Driver Ability index (from previous season drivers' championship points).

Our training label (result) is the final position gain or loss compared with the starting grid position. Using it as the label could train the ML model to understand how good the pitstop strategy is, whereas the poor result in position loss would have resulted from a poor set of pitstop strategy data input, and vice versa for good results.

While most of the source data is intuitively either an integer (e.g., starting position) or a real number (e.g., temperature), it is worth mentioning **how we transform the tyre strategy data** as input data into the ML model. According to the official rules from 2014 to 2020 seasons (which belongs to the same era with 1.6-litre Hybrid V6 Turbo Engine), there are a total of 5 types of dry tyre compound in seasons practically, and 2 types of wet tyre compound. Since it is a categorical data regarding of which type of tyre compound is being used in a pit stop strategy, one-hot-encoding method is recommended and applied initially. The problem of such representation is that we have to duplicate all 5 dry tyre and 2 wet tyre choices for at least 3 times to represent a 2-time pit stop strategy.

dry1_start	dry2_start	dry3_start	dry4_start	dry5_start	wet1_start	wet2_start
0	0	1	0	0	0	0
dry1_pit1	dry2_pit1	dry3_pit1	dry4_pit1	dry5_pit1	wet1_pit1	wet2_pit1
0	0	0	1	0	0	0
dry1_pit2	dry2_pit2	dry3_pit2	dry4_pit2	dry5_pit2	wet1_pit2	wet2_pit2
0	0	0	1	0	0	0

Representation of a 2-time pit stop strategy using one-hot-encoding.

This approach would not only cause the input layer of the ML model to be very large, which may affect the training efficiency and accuracy, but also it has not represented when (i.e., on which lap) the pitstop takes place during a race. As a result, we figure out that the relationship between dry tyres is the level of stiffness (how hard the compound is), therefore can be represented logically by a scale from 1 to 5 according to its stiffness. We keep one-hot-encoding for a separate model for wet conditions so to train a model that is dry-condition focused and efficient.

Dealing with which lap did the pitstop take place, since almost every racetrack has a different number of laps, using the lap number directly does not create a fair comparison across different tracks. Yet, using percentages to represent when the pitstop is not ideal as well, since a 1-stop and a 2-stop strategy cannot be well represented at the same time. For example, using the aforementioned A) lap number and B) percentage representation method, in a 50-lap race, representing 1) a 1-stop strategy at lap 23, and 2) a 2-stop strategy at lap 15 and lap 37, would be as follows:

A1

first_stop	second_stop
23	N.A.

B1

first_stop	second_stop
0.46	N.A.

A2

first_stop	second_stop
15	37

B2

first_stop	second_stop
0.3	0.74

It shows that the second stop could not be simply replaced by a 0 since it has numerical meaning in such representation. After several modification on representation, by considering the stiffness and tyre distance as the input factors, the result is as follows:

tyre_start	tyre_pit1	tyre_pit2	tyre_start_distance	tyre_pit1_distance	tyre_pit2_distance
3	4	4	23	27	0

1-stop strategy at lap 23, from tyre type-3 to type-4.

tyre_start	tyre_pit1	tyre_pit2	tyre_start_distance	tyre_pit1_distance	tyre_pit2_distance
3	4	4	15	22	13

2-stop strategy at lap 15 and lap 37, from tyre type-3 to type-4, and to type-4 again.

As shown on the 2 instances, we can distinguish a 1-stop and 2-stop pitstop strategy by setting “tyre_pit2_distance” to be 0 or an integer: A type-4 tyre at pit2 traveling 0 laps means it remains using type-4 after the first pit and continues to travel 0 more laps, which basically mean it did not pit in again. The second case with “tyre_pit2_distance” of 13 laps means the car made a second pitstop and traveled 13 laps on that tyre compound.

Preprocessing of the dataset

Source of data

The majority of data relies on FIA Championships Results & Statistics website at https://fiaresultsandstatistics.motorsportstats.com/results. For each single race event, under “Classification” and “Session Facts”, we can extract each racer’s race number, initial position, final position, and a summary of the pitstop strategies.

Fastest Laps Lap Chart Max Speeds Lap Times Pit Stop Summary

No.	Driver	Team	Lap	Stop
3	Daniel Ricciardo	Renault DP World F1 Team	39	1
4	Lando Norris	McLaren F1 Team	10	1
5	Sebastian Vettel	Scuderia Ferrari	35	1
6	Nicholas Latifi	Williams Racing	11	1
6	Nicholas Latifi	Williams Racing	35	2
7	Kimi Räikkönen	Alfa Romeo Racing ORLEN	10	1
10	Pierre Gasly	Scuderia AlphaTauri Honda	10	1
16	Charles Leclerc	Scuderia Ferrari	22	1
18	Lance Stroll	BWT Racing Point F1 Team	10	1
20	Kevin Magnussen	Haas F1 Team	31	1
20	Kevin Magnussen	Haas F1 Team	47	2
23	Alexander Albon	Aston Martin Red Bull Racing	10	1
26	Daniil Kvyat	Scuderia AlphaTauri Honda	10	1

Dataset from FIA, Abu Dhabi GP 2020 Pit Stop Summary.

However, we would only have the information of the lap number when the racer goes into the pitstop, and whether it is a first, second or third pitstop. Therefore, we have to rely on our next data source RaceFans at https://www.racefans.net/category/20120-f1-season/2020-f1-race-data/.

2020 Abu Dhabi Grand Prix tyre strategies

The tyre strategies for each driver:

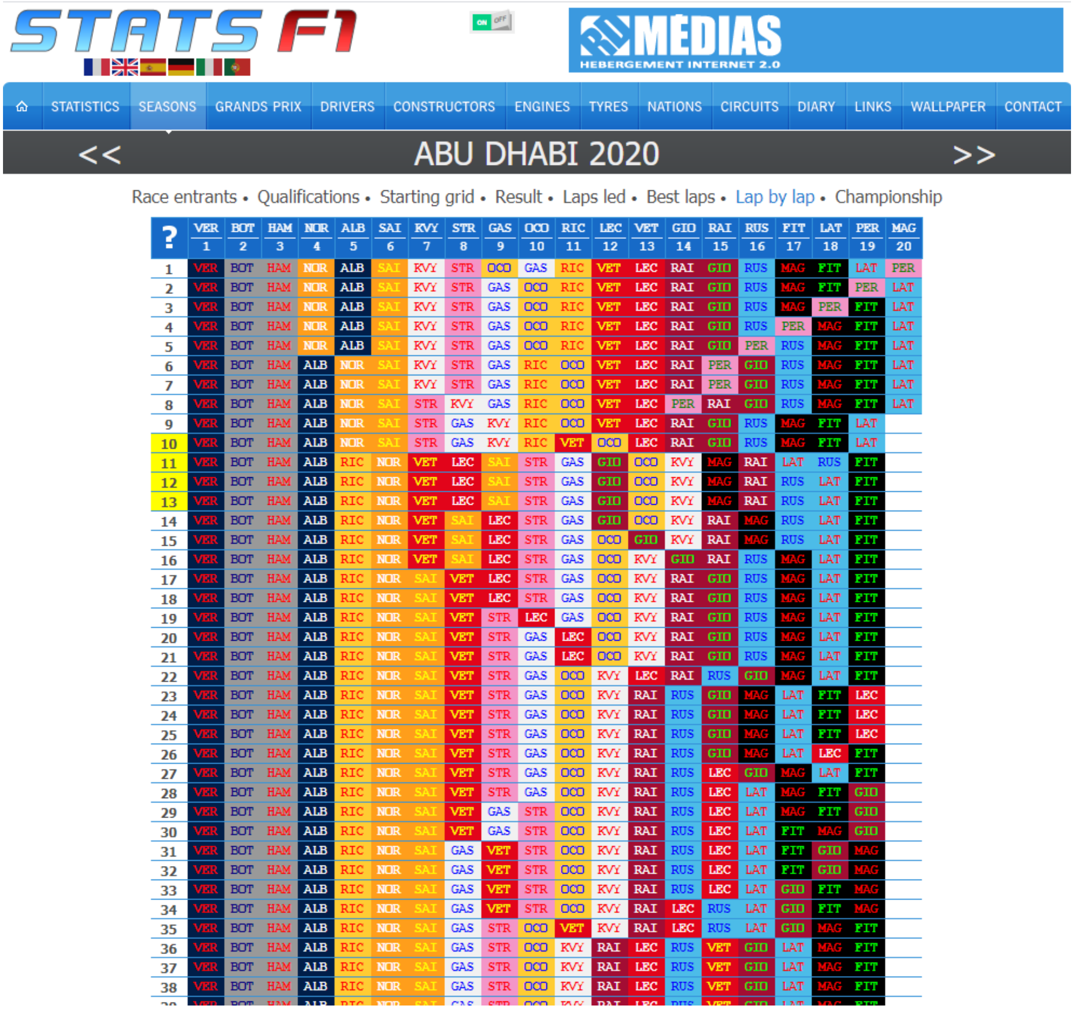
	Stint 1	Stint 2	Stint 3	Stint 4
Max Verstappen	C4 (10)	C3 (45)		
Valtteri Bottas	C4 (10)	C3 (45)		
Lewis Hamilton	C4 (10)	C3 (45)		
Alexander Albon	C5 (10)	C3 (45)		
Lando Norris	C5 (10)	C3 (45)		
Carlos Sainz Jnr	C4 (10)	C3 (45)		
Daniel Ricciardo	C3 (39)	C4 (16)		
Pierre Gasly	C5 (10)	C3 (45)		
Esteban Ocon	C4 (10)	C3 (45)		
Lance Stroll	C5 (10)	C3 (45)		
Daniil Kvyat	C5 (10)	C3 (44)		
Kimi Raikkonen	C4 (10)	C3 (44)		
Charles Leclerc	C4 (22)	C3 (32)		
Sebastian Vettel	C3 (35)	C4 (19)		
George Russell	C4 (10)	C3 (44)		
Antonio Giovinazzi	C4 (27)	C3 (27)		
Nicholas Latifi	C4 (11)	C3 (24)	C4 (19)	
Kevin Magnussen	C3 (31)	C4 (16)	C4 (7)	
Pietro Fittipaldi	C4 (10)	C3 (24)	C4 (14)	C5 (5)
Sergio Perez	C3 (8)			

Dataset from RaceFans, Abu Dhabi GP 2020 Pit Stop strategy.

From this website, we can extract data such as which tyre compound was used, and what was the traveled distance of each tyre set (in terms of number of laps). For example, the first row describes Max Verstappen (Racer No. 33) used C4 tyre compound to travel 10 laps, and changed to C3 tyre compound to travel another 45 laps to finish the race.

In order to check if the racer made the pitstop under a safety car condition (explained at Model 2), we take advantage of StatsF1 at https://www.statsf1.com/en/2020.aspx. Inside the “lap by lap” chart, the laps with safety car / virtual safety car are highlighted in yellow.

Name: Chan Ming Chung
UID: 3035373169

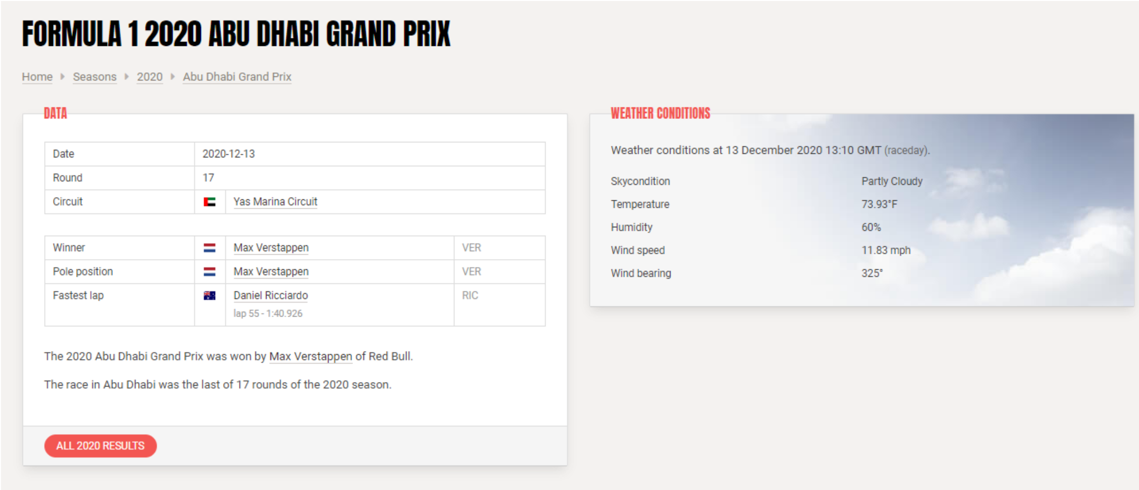


?	VER 1	BOT 2	HAM 3	NOR 4	ALB 5	SAI 6	KVY 7	STR 8	GAS 9	OCO 10	RIC 11	LEC 12	VET 13	GIO 14	RAI 15	RUS 16	FIT 17	LAT 18	PER 19	MAG 20
1	VER	BOT	HAM	NOR	ALB	SAI	KVY	STR	OCO	GAS	RIC	VET	LEC	RAI	GIO	RUS	MAG	FIT	LAT	PER
2	VER	BOT	HAM	NOR	ALB	SAI	KVY	STR	GAS	OCO	RIC	VET	LEC	RAI	GIO	RUS	MAG	FIT	PER	LAT
3	VER	BOT	HAM	NOR	ALB	SAI	KVY	STR	GAS	OCO	RIC	VET	LEC	RAI	GIO	RUS	MAG	PER	FIT	LAT
4	VER	BOT	HAM	NOR	ALB	SAI	KVY	STR	GAS	OCO	RIC	VET	LEC	RAI	GIO	RUS	PER	MAG	FIT	LAT
5	VER	BOT	HAM	NOR	ALB	SAI	KVY	STR	GAS	OCO	RIC	VET	LEC	RAI	GIO	PER	RUS	MAG	FIT	LAT
6	VER	BOT	HAM	ALB	NOR	SAI	KVY	STR	GAS	RIC	OCO	VET	LEC	RAI	PER	GIO	RUS	MAG	FIT	LAT
7	VER	BOT	HAM	ALB	NOR	SAI	KVY	STR	GAS	RIC	OCO	VET	LEC	RAI	PER	GIO	RUS	MAG	FIT	LAT
8	VER	BOT	HAM	ALB	NOR	SAI	STR	KVY	GAS	RIC	OCO	VET	LEC	PER	RAI	GIO	RUS	MAG	FIT	LAT
9	VER	BOT	HAM	ALB	NOR	SAI	STR	GAS	KVY	RIC	OCO	VET	LEC	RAI	GIO	RUS	MAG	FIT	LAT	
10	VER	BOT	HAM	ALB	NOR	SAI	STR	GAS	KVY	RIC	VET	OCO	LEC	RAI	GIO	RUS	MAG	FIT	LAT	
11	VER	BOT	HAM	ALB	RIC	NOR	VET	LEC	SAI	STR	GAS	GIO	OCO	KVY	MAG	RAI	LAT	RUS	FIT	
12	VER	BOT	HAM	ALB	RIC	NOR	VET	LEC	SAI	STR	GAS	GIO	OCO	KVY	MAG	RAI	RUS	LAT	FIT	
13	VER	BOT	HAM	ALB	RIC	NOR	VET	LEC	SAI	STR	GAS	GIO	OCO	KVY	MAG	RAI	RUS	LAT	FIT	
14	VER	BOT	HAM	ALB	RIC	NOR	VET	SAI	LEC	STR	GAS	GIO	OCO	KVY	RAI	MAG	RUS	LAT	FIT	
15	VER	BOT	HAM	ALB	RIC	NOR	VET	SAI	LEC	STR	GAS	OCO	GIO	KVY	RAI	MAG	RUS	LAT	FIT	
16	VER	BOT	HAM	ALB	RIC	NOR	VET	SAI	LEC	STR	GAS	OCO	KVY	GIO	RAI	RUS	MAG	LAT	FIT	
17	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	LEC	STR	GAS	OCO	KVY	RAI	GIO	RUS	MAG	LAT	FIT	
18	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	LEC	STR	GAS	OCO	KVY	RAI	GIO	RUS	MAG	LAT	FIT	
19	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	STR	LEC	GAS	OCO	KVY	RAI	GIO	RUS	MAG	LAT	FIT	
20	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	STR	GAS	LEC	OCO	KVY	RAI	GIO	RUS	MAG	LAT	FIT	
21	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	STR	GAS	LEC	OCO	KVY	RAI	GIO	RUS	MAG	LAT	FIT	
22	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	STR	GAS	OCO	KVY	LEC	RAI	RUS	GIO	MAG	LAT	FIT	
23	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	STR	GAS	OCO	KVY	RAI	RUS	GIO	MAG	LAT	FIT	LEC	
24	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	STR	GAS	OCO	KVY	RAI	RUS	GIO	MAG	LAT	FIT	LEC	
25	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	STR	GAS	OCO	KVY	RAI	RUS	GIO	MAG	LAT	FIT	LEC	
26	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	STR	GAS	OCO	KVY	RAI	RUS	GIO	MAG	LAT	LEC	FIT	
27	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	STR	GAS	OCO	KVY	RAI	RUS	LEC	GIO	MAG	LAT	FIT	
28	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	STR	GAS	OCO	KVY	RAI	RUS	LEC	LAT	MAG	FIT	GIO	
29	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	GAS	STR	OCO	KVY	RAI	RUS	LEC	LAT	MAG	FIT	GIO	
30	VER	BOT	HAM	ALB	RIC	NOR	SAI	VET	GAS	STR	OCO	KVY	RAI	RUS	LEC	LAT	FIT	MAG	GIO	
31	VER	BOT	HAM	ALB	RIC	NOR	SAI	GAS	VET	STR	OCO	KVY	RAI	RUS	LEC	LAT	FIT	GIO	MAG	
32	VER	BOT	HAM	ALB	RIC	NOR	SAI	GAS	VET	STR	OCO	KVY	RAI	RUS	LEC	LAT	FIT	GIO	MAG	
33	VER	BOT	HAM	ALB	RIC	NOR	SAI	GAS	VET	STR	OCO	KVY	RAI	RUS	LEC	LAT	GIO	FIT	MAG	
34	VER	BOT	HAM	ALB	RIC	NOR	SAI	GAS	VET	STR	OCO	KVY	RAI	LEC	RUS	LAT	GIO	FIT	MAG	
35	VER	BOT	HAM	ALB	RIC	NOR	SAI	GAS	STR	OCO	VET	KVY	RAI	LEC	RUS	LAT	GIO	MAG	FIT	
36	VER	BOT	HAM	ALB	RIC	NOR	SAI	GAS	STR	OCO	KVY	RAI	LEC	RUS	VET	GIO	LAT	MAG	FIT	
37	VER	BOT	HAM	ALB	RIC	NOR	SAI	GAS	STR	OCO	KVY	RAI	LEC	RUS	VET	GIO	LAT	MAG	FIT	
38	VER	BOT	HAM	ALB	RIC	NOR	SAI	GAS	STR	OCO	KVY	RAI	LEC	RUS	VET	GIO	LAT	MAG	FIT	

Dataset from StatsF1, lap by lap chart.

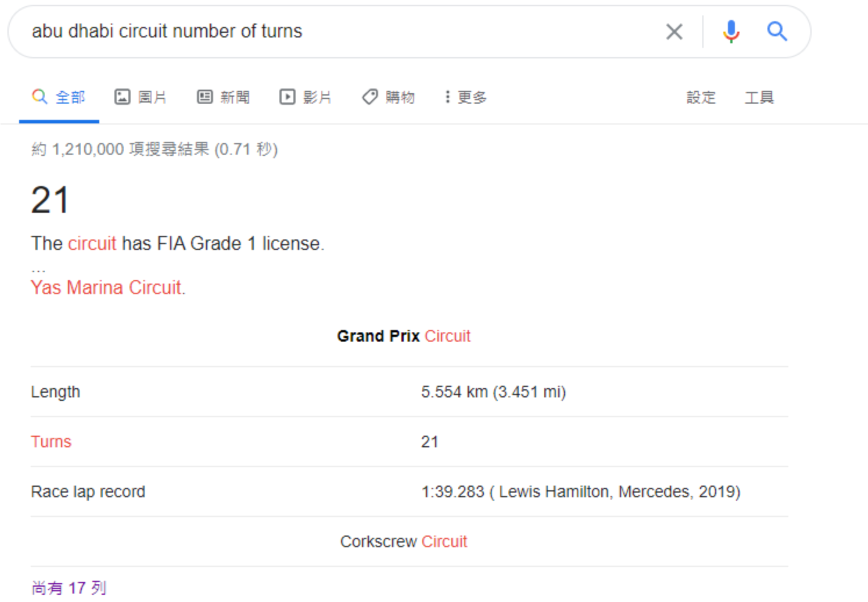
Locating the laps, we cross-check our first FIA dataset and review if the racer went into the pitstop during lap 10 - lap13. We count the number of times a racer chose to make a pitstop under the safety car condition in each race, and store the result under the column "sc", which is short for safety car.

Regarding the environment of the race event, we extract weather information from Racing Statistics at https://www.racing-statistics.com/en. It includes not only the temperature and humidity, but also wind speed, wind bearing, and a general description of the sky condition, which we may not use in training the model.



Dataset from Racing Statistics, Weather conditions.

There are approximately 20 races in each season, which means there are about 20 different tracks with various track design. An intuitive way is to use one-hot-encoding to categorize each race circuit. However, there are minor changes made in race circuits throughout 2014-2019, which includes some introductions of new tracks, deleting old tracks. In 2020 and 2021 the world faced the unprecedented COVID pandemic situation, which forced F1 to design a 17-race new racing calendars. Half of the race circuits are new to F1, and teams have no prior experience nor data collected from these circuits. The use of one-hot-encoding will therefore be useless when the model tries to predict a new race circuit event. As a result, we extract characteristics from each race circuit and generalize how the model should understand the track difference. Altitude change, number of turns, total racing laps, and total track distance are publicly known data and can be extracted from F1 website or any third-party websites. These data hold comparatively consistent throughout each season, therefore the data columns are batch processed in the Python program instead of manually inserting into each excel file.



Sample of getting number of turns in Abu Dhabi Yas Marina Circuit

Finally, we should assume that every racer and every team perform differently, because some racers are more aggressive and experienced, or some teams have better engines and aerodynamics, which heavily constitutes to race performance. We use previous year's Drivers Championship points and Constructors (Team) championship points to predict each racers and each team's ability in the form of a referencing index score. The data can be extracted from FIA website as well.

Name: Chan Ming Chung
UID: 3035373169

EVENT INFO | CLASSIFICATION | SESSION FACTS | STANDINGS | STATS

Driver Standings | Team Standings

Pos.	Driver	Points	AUS	BHR	CHN	AZE	ESP	MCO	CAN	FRA	AUT	GBR	DEU	HUN	BEL	ITA	SGP	RUS	JPN	MEX	USA	BRA
1	Lewis Hamilton	413	18	25	25	18	26	25	25	25	10	26	2	25	18	16	12	26	16	25	18	6
2	Valtteri Bottas	326	26	18	18	25	18	15	13	18	15	18	0	4	15	18	10	18	25	15	25	0
3	Max Verstappen	278	15	12	12	12	15	12	10	12	26	10	26	19	0	4	15	12	0	8	15	25
4	Charles Leclerc	264	10	16	10	11	10	0	15	15	18	15	0	12	25	25	18	15	8	13	13	0
5	Sebastian Vettel	240	12	10	15	15	12	18	18	11	12	0	18	15	13	0	25	0	18	18	0	0
6	Carlos Sainz	96	DNF	0	0	6	4	8	0	8	4	8	10	10	0	0	0	8	10	0	4	15
7	Pierre Gasly	95	0	4	9	0	8	11	4	1	6	12	0	8	2	0	4	0	6	2	0	18
8	Alexander Albon	92	0	2	1	0	0	4	0	0	0	0	8	1	10	8	8	10	12	10	10	0
9	Daniel Ricciardo	54	DNF	0	6	0	0	2	8	0	0	6	0	0	0	12	0	0	0	4	8	8
10	Sergio Pérez	52	0	1	4	8	0	0	0	0	0	0	0	0	8	6	0	6	4	6	1	2
11	Lando Norris	49	0	8	0	4	0	0	0	2	8	0	0	2	0	1	6	4	0	0	6	4
12	Kimi Räikkönen	43	4	6	2	1	0	0	0	6	2	4	0	6	0	0	0	0	0	0	0	12
13	Daniil Kvyat	37	1	0	0	0	2	6	1	0	0	2	15	0	6	0	0	0	1	0	0	1
14	Nico Hülkenberg	37	6	0	0	0	0	0	6	4	0	1	0	0	4	10	2	1	0	1	2	0
15	Lance Stroll	21	2	0	0	2	0	0	2	0	0	0	12	0	1	0	0	0	2	0	0	0
16	Kevin Magnussen	20	8	0	0	0	6	0	0	0	0	0	4	0	0	0	0	2	0	0	0	0
17	Antonio Giovinazzi	14	0	0	0	0	0	0	0	0	1	0	0	0	0	2	1	0	0	0	0	10
18	Romain Grosjean	8	DNF	DNF	0	0	1	1	0	0	0	0	6	0	0	0	0	0	0	0	0	0
19	Robert Kubica	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
20	George Russell		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Dataset from FIA, Driver’s Championship 2019.

EVENT INFO | CLASSIFICATION | SESSION FACTS | STANDINGS | STATS

Driver Standings | Team Standings

Pos.	Team	Points	AUS	BHR	CHN	AZE	ESP	MCO	CAN	FRA	AUT	GBR	DEU	HUN	BEL	ITA	SGP	RUS	JPN	MEX	U
1	Mercedes-AMG Petronas…	739	44	43	43	43	44	40	38	43	25	44	2	29	33	34	22	44	41	40	[illegible]
2	Scuderia Ferrari Mission …	504	22	26	25	26	22	18	33	26	30	15	18	27	38	25	43	15	26	31	[illegible]
3	Aston Martin Red Bull Ra…	417	15	16	21	12	23	23	14	13	32	22	26	27	10	12	23	22	12	18	[illegible]
4	McLaren F1 Team	145	0	8	0	10	4	8	0	10	12	8	10	12	0	1	6	12	10	0	[illegible]
5	Renault F1 Team	91	6	0	6	0	0	2	14	4	0	7	0	0	4	22	2	1	0	5	[illegible]
6	Red Bull Toro Rosso Honda	85	1	2	1	0	2	10	1	0	0	2	23	1	8	0	4	0	7	2	
7	SportPesa Racing Point F…	73	2	1	4	10	0	0	2	0	0	0	12	0	9	6	0	6	6	6	
8	Alfa Romeo Racing	57	4	6	2	1	0	0	0	6	3	4	0	6	0	2	1	0	0	0	
9	Haas F1 Team	28	8	0	0	0	7	1	0	0	0	0	10	0	0	0	0	2	0	0	
10	ROKiT Williams Racing	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	

Dataset from FIA, Constructor’s Championship 2019.

Name: Chan Ming Chung
UID: 3035373169

Data cleaning and preprocessing

No.	initial_pos	final_pos	no_of_pits	tyre_grid	tyre_1	tyre_2	tyre_grid_distance	tyre_1_distance	tyre_2_distance	sc	final_pos_gained	final_points
3	11	7	1	3	4	4	39	16	0	0	4	6
4	4	5	1	5	3	3	10	45	0	1	-1	10
5	13	14	1	3	4	4	35	19	0	0	-1	0
6	18	17	2	4	3	4	11	24	19	1	1	0
7	15	12	1	4	3	3	10	44	0	1	3	0
10	9	8	1	5	3	3	10	45	0	1	1	4
11	19	DNF								0		0
16	12	13	1	4	3	3	22	32	0	0	-1	0
18	8	10	1	5	3	3	10	45	0	1	-2	1
20	20	18	2	3	4	4	31	16	7	0	2	0
23	5	4	1	5	3	3	10	45	0	1	1	12
26	7	11	1	5	3	3	10	44	0	1	-4	0
31	10	9	1	4	3	3	10	45	0	1	1	2
33	1	1	1	4	3	3	10	45	0	1	0	25
44	3	3	1	4	3	3	10	45	0	1	0	15
51	17	19	3							1	-2	0
55	6	6	1	4	3	3	10	45	0	1	0	8
63	16	15	1	4	3	3	10	44	0	1	1	0
77	2	2	1	4	3	3	10	45	0	1	0	18
99	14	16	1	4	3	3	27	27	0	0	-2	0

Sample dataset from Abu Dhabi GP 2019.

Since there are DNF (Did Not Finish) or DSQ (Disqualified) records, we will drop these rows because the pit stop strategy does not fully cause the race retirement, where there could be a mixture of engine issues, tyre malfunction, car damage, or even racing incidents such as car crashing to each other or the wall. Racers that made more than 2 pitstops would also be dropped from the dataset, since the optimal number of pitstop without any accidents lies between 1 and 2 pitstops in the modern era of F1.

No.	initial_pos	final_pos	no_of_pits	tyre_grid	tyre_1	tyre_2	tyre_grid_distance	tyre_1_distance	tyre_2_distance	sc	final_pos_gained	final_points
3	13	DNF	1	W	I	I	3	61	0	1		0
4	19	DNF	2	W	I	4	6	19	39	0		0
5	20	2	5	W	I	4	2	21	0	2	18	18
7	5	12	4	W	I	3	3	23	0	2	-7	0
8	6	7	5	W	I	4	3	23	0	3	-1	6
10	4	14	4	W	I	4	3	23	0	2	-10	0
11	8	DNF	0							0		0
16	10	DNF	3	W	I	I	3	12	0	2		0
18	15	4	5	W	I	3	7	18	0	3	11	12
20	12	8	6	W	I	4	8	13	0	4	4	4
23	16	6	4	W	I	4	2	25	0	1	10	8
26	14	3	4	W	I	4	3	22	0	3	11	15
27	9	DNF	3	W	I	I	3	12	0	3		0
33	2	1	5	W	I	3	3	22	0	3	1	25
44	1	9	6	W	I	4	3	25	0	4	-8	2
55	7	5	3	W	I	I	3	24	0	1	2	10
63	17	11	5	W	I	I	7	17	0	2	6	0
77	3	DNF	4	W	I	3	3	23	0	2		0
88	18	10	5	W	I	I	7	16	0	2	8	1
99	11	13	4	W	I	3	3	23	0	2	-2	0

Sample dataset from German GP 2019.

Since there are rainy races such as German GP 2019, in which drivers may choose from Wet compound (denoted as W) or Intermediate compound (denoted as I), or even switch back to dry compounds when the track is acceptably dry enough, these data should not be trained together with dry condition races. Therefore, we removed these records for our model.

Combining all aforementioned data, we finally create a huge dataset that has 18 columns.

initial_pos	no_of_pits	tyre_grid	tyre_1	tyre_2	tyre_grid_distance	tyre_1_distance	tyre_2_distance	sc	Temperature	Humidity	Altitude	Turns	RaceDistance	TrackLength	TeamAbility	DriverAbility	final_pos_gained
18	1	3	4	4	35	19	0	0	90.49	0.38	10.7	21	55	5.554	62	12	4
5	1	4	3	3	33	22	0	0	90.49	0.38	10.7	21	55	5.554	419	170	1
3	1	4	3	3	15	40	0	0	90.49	0.38	10.7	21	55	5.554	571	320	1
7	1	5	3	3	7	47	0	1	90.49	0.38	10.7	21	55	5.554	93	37	-2
14	1	4	3	3	26	29	0	0	90.49	0.38	10.7	21	55	5.554	0	62	6
15	1	4	3	3	27	27	0	0	90.49	0.38	10.7	21	55	5.554	62	50	4
20	1	3	4	4	40	14	0	0	90.49	0.38	10.7	21	55	5.554	7	6	7
13	1	3	4	4	41	13	0	0	90.49	0.38	10.7	21	55	5.554	93	56	3
16	1	4	3	3	1	53	0	1	90.49	0.38	10.7	21	55	5.554	33	4	4
6	1	5	3	3	17	38	0	0	90.49	0.38	10.7	21	55	5.554	419	249	3
19	1	3	4	4	35	19	0	0	90.49	0.38	10.7	21	55	5.554	7	1	4
11	1	4	3	3	37	18	0	0	90.49	0.38	10.7	21	55	5.554	122	53	5
2	2	4	3	4	16	24	15	0	90.49	0.38	10.7	21	55	5.554	655	247	-3
11	1	4	3	3	25	33	0	1	75.07	0.35	2.6	16	58	5.303	62	12	2
8	1	3	2	2	26	32	0	1	75.07	0.35	2.6	16	58	5.303	419	170	4
3	1	4	2	2	26	32	0	1	75.07	0.35	2.6	16	58	5.303	571	320	2
2	1	4	2	2	18	40	0	0	75.07	0.35	2.6	16	58	5.303	571	251	-1
12	1	4	2	2	24	34	0	0	75.07	0.35	2.6	16	58	5.303	52	62	1
10	1	4	2	2	26	32	0	1	75.07	0.35	2.6	16	58	5.303	62	50	5
18	2	3	2	4	20	7	31	1	75.07	0.35	2.6	16	58	5.303	48	39	5
13	2	3	2	4	25	4	29	1	75.07	0.35	2.6	16	58	5.303	7	6	-1
14	1	4	2	2	23	35	0	0	75.07	0.35	2.6	16	58	5.303	52	49	2
4	1	3	2	2	21	37	0	0	75.07	0.35	2.6	16	58	5.303	419	249	-2
1	1	4	2	2	19	39	0	0	75.07	0.35	2.6	16	58	5.303	655	408	-1
9	1	4	2	2	22	36	0	0	75.07	0.35	2.6	16	58	5.303	122	53	-1
15	1	4	3	3	25	33	0	1	75.07	0.35	2.6	16	58	5.303	655	247	7
14	2	3	2	2	1	39	25	0	71.62	0.46	63.5	8	71	4.318	62	12	-1
7	1	4	2	2	15	56	0	1	71.62	0.46	63.5	8	71	4.318	571	320	4
7	1	4	2	2	15	55	0	1	71.62	0.46	63.5	8	71	4.318	93	37	3
16	1	2	3	3	45	25	0	0	71.62	0.46	63.5	8	71	4.318	48	9	6

30-row sample of the final dataset to be fed into the model.

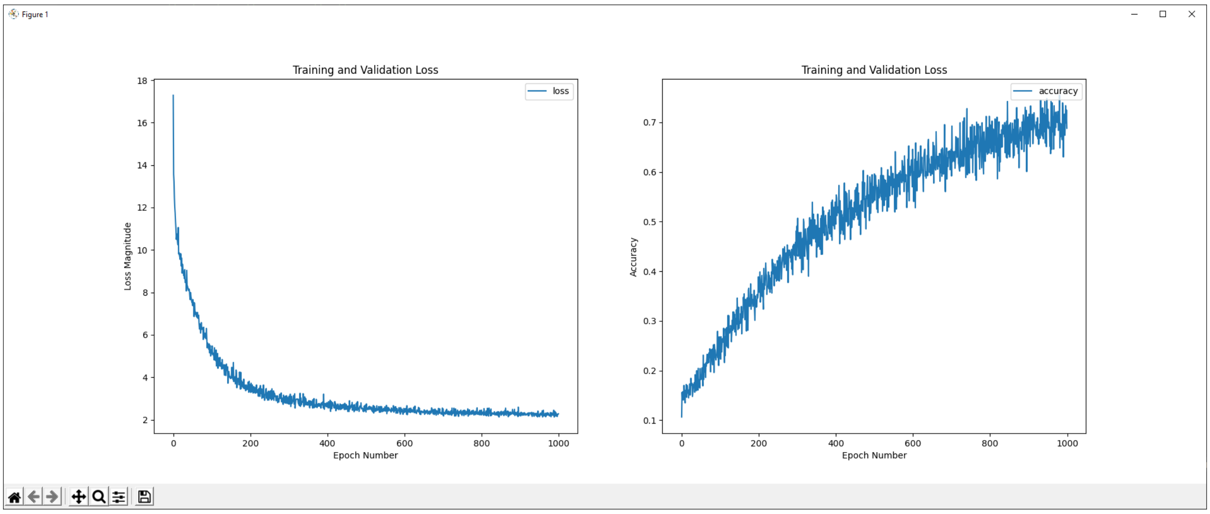
Model design evolutions

Version 1: 3 Dense Layers model, predicting on "final position changes"

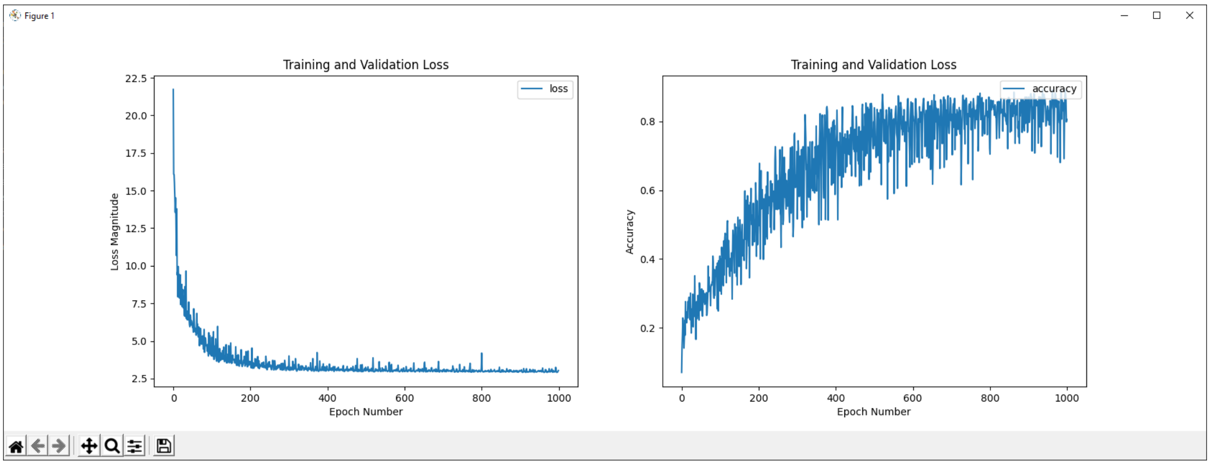
The first design of the DNN model consisted of 3 Dense layers with 256 neurons at the first layer. Linear activation function was used, and a soft-accuracy metric was adopted by taking the mean of the difference between predicted final position changes and the real-life position changes. We adopted mean-squared-error for the loss function, Selu for activation, and rmsprop for the optimizer. By then, there were only 10 columns for all input data, since we ignored how driver's ability, car's ability, and the characteristics of the race track and pit stop decisions under the safety car condition. Running 1000 epochs when training the model could result in a maximum 19% of accuracy. If we only include the top 3 racing teams, namely Mercedes Benz, Scuderia Ferrari, and Red Bull racing, the maximum accuracy is still 35%.

Version 2: More sophisticated 8-layer DNN model

The next design introduced a more sophisticated 8-layer DNN that handles the same 10-column inputs. The model performed in increasing accuracy and decreasing loss over the number of epochs.

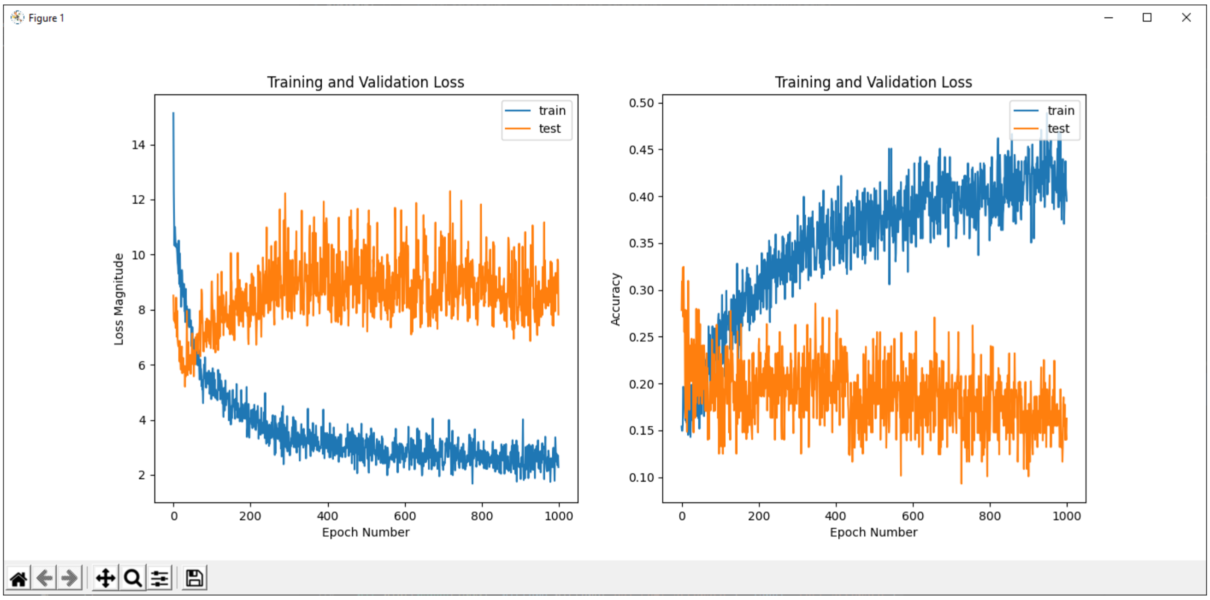


Loss and accuracy for data from 2017-2019.



Loss and accuracy for data from 2017-2019, with Top 3 Teams only.

For the fear of overfitting, we created an 80/20 split to the dataset for training data and testing data so to investigate the "too-good" accuracy results. We plotted a graph of both training data and testing data accuracy and loss against the number of epochs, and the results are as follows:



Loss and accuracy for training & testing data from 2017-2019

As the number of epochs increases, although an increase in accuracy in the training data curve is observed, the testing data curve poorly improve in accuracy; Similar situation occurs in loss where testing loss is much higher than training loss, and not improving in the same speed as that of the training curve. This implies clearly that the model was in an overfitting situation, where the model tried to remember data patterns instead of generalizing knowledge from the training dataset.

To solve the overfitting problem, several strategies could be applied:

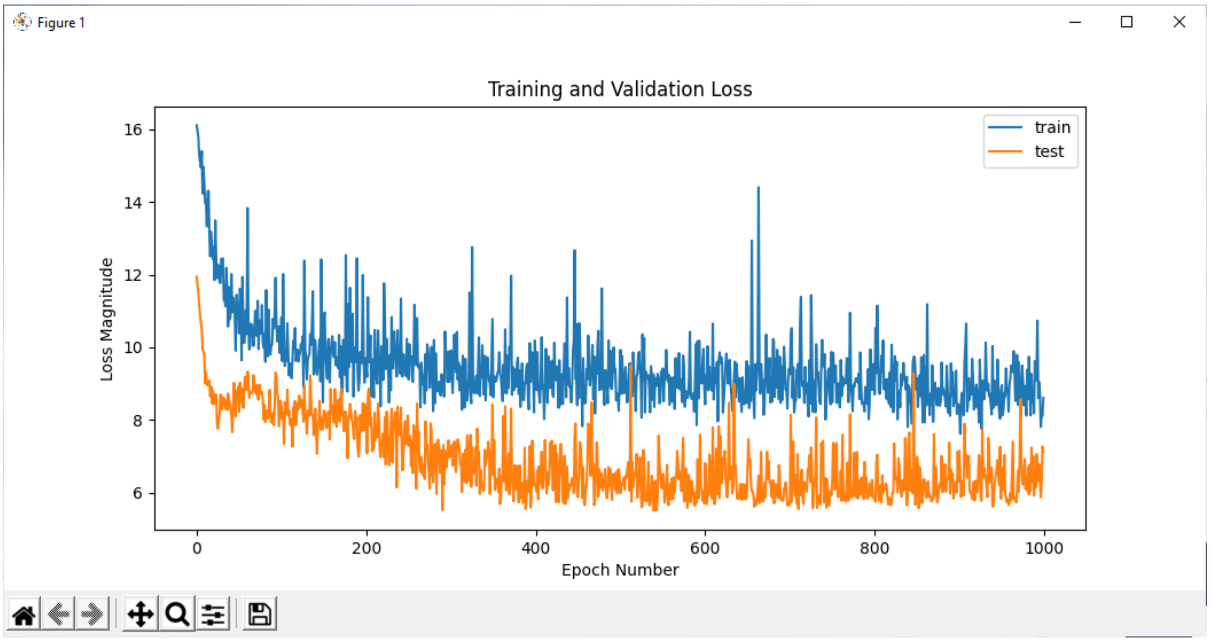
1. Simplifying the model.
2. Apply normalization layer and regularization.
3. Add Dropout Layers.
4. Early Stopping.
5. Feature extraction & feature selection.
6. Refine custom metric function.

7. Incorporate data from 2014-2017 as well.
8. Include racer related data.
9. Include race car-related data.

Where the first 1 to 5 strategies relate to model architecture design; The 6th strategy refines the definition of accuracy; Strategy number 7 to 9 correlate to the vertical expansion of dataset by introducing more race seasons, and the horizontal expansion by exploring new data columns that could be helpful in the prediction.

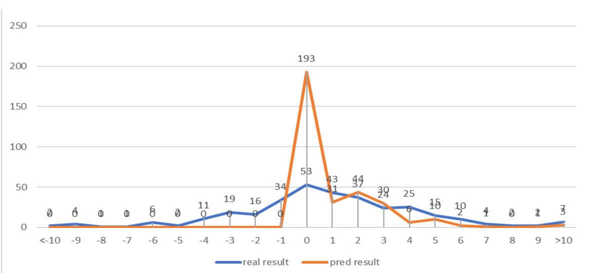
Version 3: Model without overfitting, but an imbalance in dataset result labels

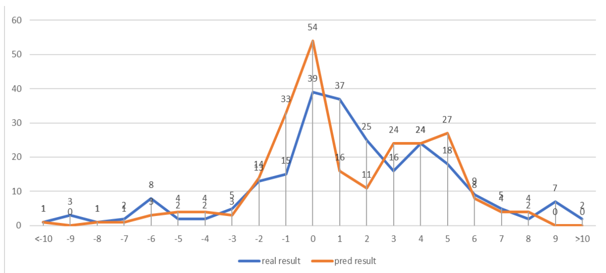
To simplify the model while respecting the complexity of the prediction model due to 17 input labels of different data, we stayed with the 8-layer DNN design while reducing the number of neutrons in each layer, ranging from 256 neutrons in the beginning layers to 16 neutrons towards the last layer. We used Tensorflow Keras Normalization function to normalize all our input labels, which predominantly constituted to the model improvement from the previous overfitting problem after careful investigation. The input layer was handled by the normalization function, and a dropout layer was applied in between each Dense layer, while L2 regularizer was also added within each Dense layer for the sake of overfitting prevention. We also incorporated data about racers and constructor teams according to the methodology mentioned at the “preprocessing of dataset section” and expand our dataset vertically to include up to the 2016 season. Judging by the effectiveness of the fix so far, we were satisfied with the improvement in both testing and training data performance.



Loss of training and testing data, data from 2017-2020 seasons

As the result shown above, a weird situation of testing data performing better than the training data was observed. The reason for such a situation was due to the fact that the model found a good-enough average result label that minimized the overall loss, or the model simply predicted the result with the highest frequency in the dataset. After investigation on the issue, we found out that the dataset consisted of mainly zeroes as the result label of “final position gain or loss”. The result labels concentrated at around zero, with very few data rows at the range of -2 to -10 and +4 to +10, which contributed to the tendency of the model in predicting zero or nearby results.





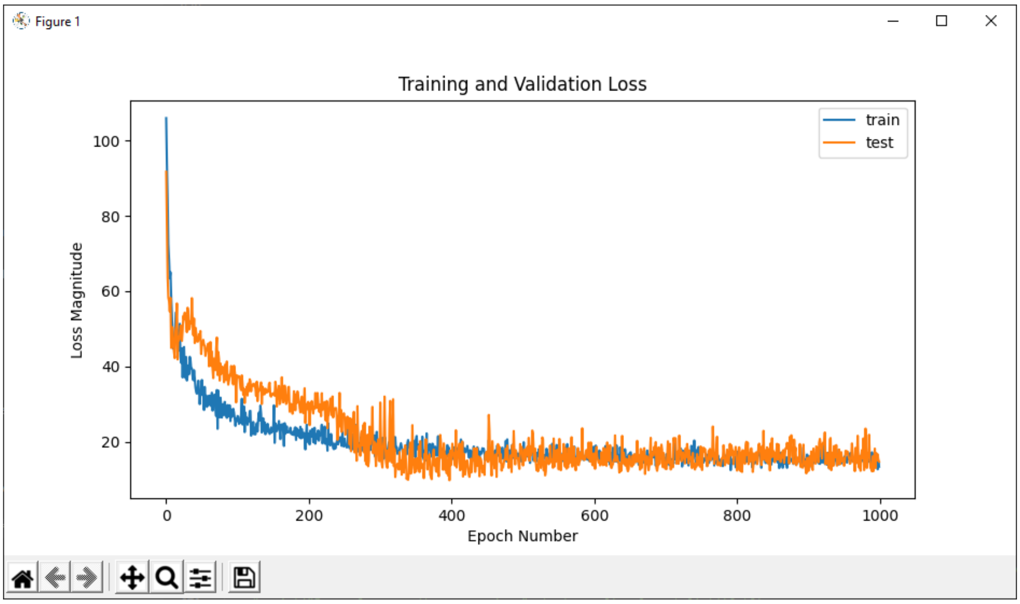
With 2016 | Without 2016

Distribution of real and predicted results, on all 2019 and 2020 season races

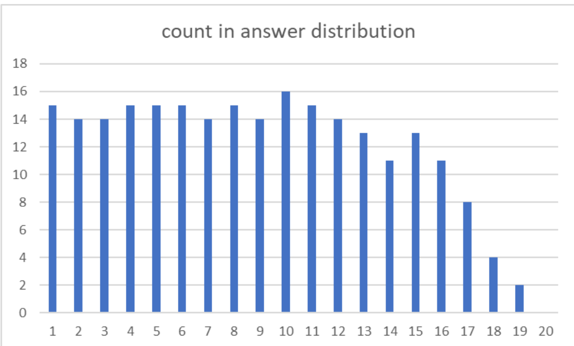
Name: Chan Ming Chung
UID: 3035373169

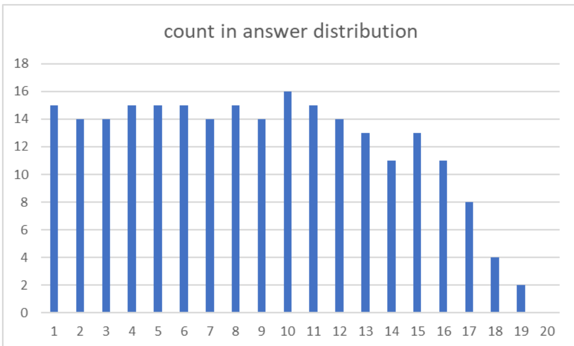
Version 4: Balanced dataset with "final position" as the result label

In each race, the final position distribution is guaranteed to be 1 racer for each position, which automatically makes "final position" the best solution to the problem of imbalance dataset result label. Adopting this model design, we ran some predictions, and the result is as follows:

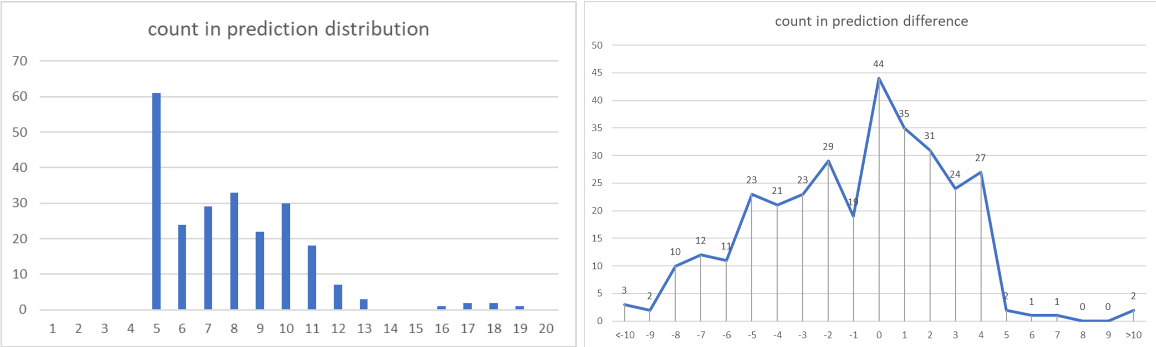


MSE Loss of training and testing data

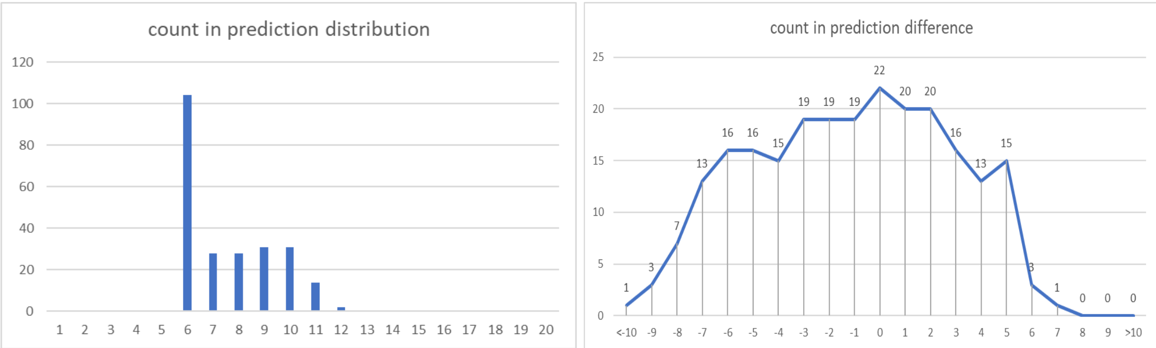




Distribution of result label "final position" in all 2019 and 2020 season races



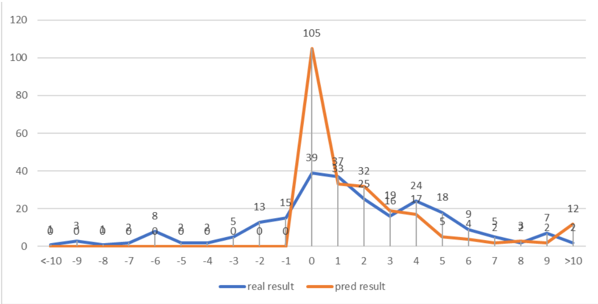
Count in prediction & Count in answer-prediction difference on all 2019 season races

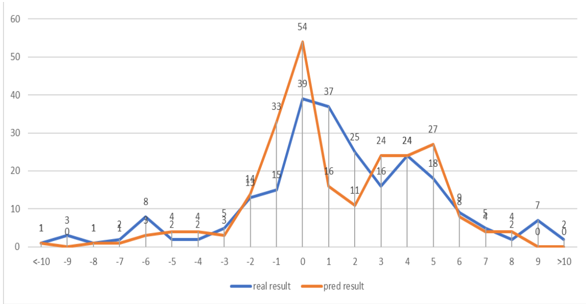


Count in prediction & Count in answer-prediction difference on all 2020 season races

As shown from the results, the model refused strongly in predicting non-average results, while only predicting from 5th to 11th position for the 2019 season testing dataset, and from 6th to 11th position for the 2020 season. It echoed the problem mentioned at Version 3: Predicting a good-enough average result so to minimize the overall loss.

Attempting to compare the difference in prediction performance of Version 3 and 4, we pondered on whether including early seasons would be beneficial to the final predictions, to see if we should include the 2014 and 2015 season results as well. There would be a chance where and new pitstop strategies and philosophies were more beneficial to the recent races, making the older seasons' tactics obsolete. As a result, we included and removed the 2016 season, which was the earliest season from our dataset, and compared the difference in prediction result. The results are as follows:





With 2016 | Without 2016

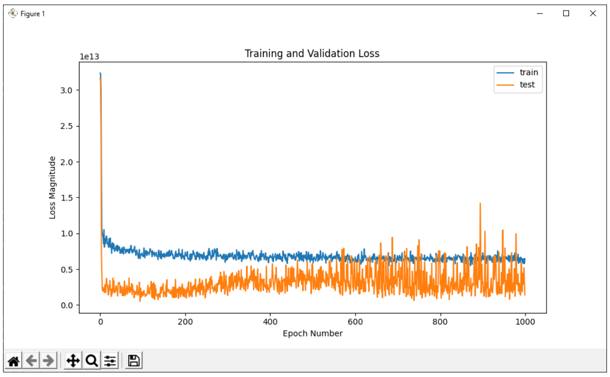
Answer and prediction results of models trained with & without the 2016 season

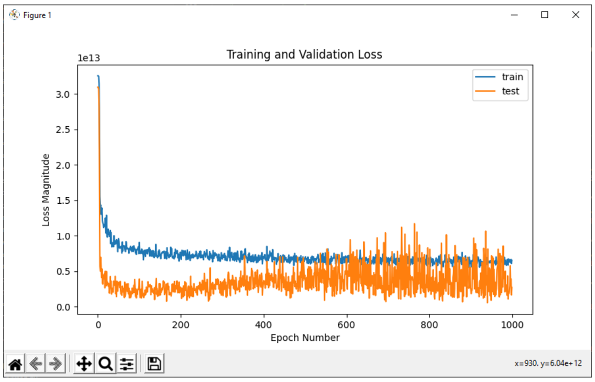
The graph suggested abnormality about including the 2016 season, where the total number of predictions results on “zero position gain” doubled, and from few predictions on negative position gain (position loss) to absolutely no negative position gain if we include the 2016 season. Considering the situation, either a new data column should be used as result label or a customized loss function that could force the model in predicting a wider range of prediction results should be introduced to improve the model.

Version 5: Exaggerating result label differences by using “Total race time” as result labels

From the FIA official website, we are able to extract the data “total race time” of each racer for each race. Since the total time used to complete the race for every racer could vary from one hundredth of a second (0.01 seconds) to 2 minutes or more, it could possibly solve the problem of over-concentration to average results.

Name: Chan Ming Chung
UID: 3035373169



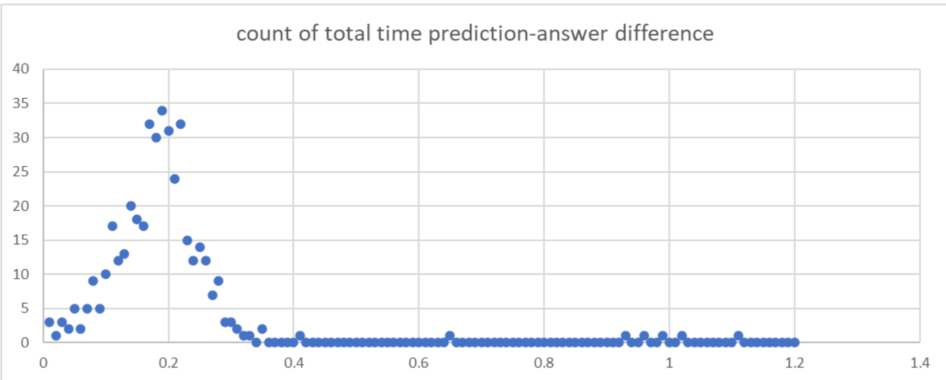


2016-2020 | 2017-2020

Test loss: 21.72 minute   Test loss: 31.07 minutes

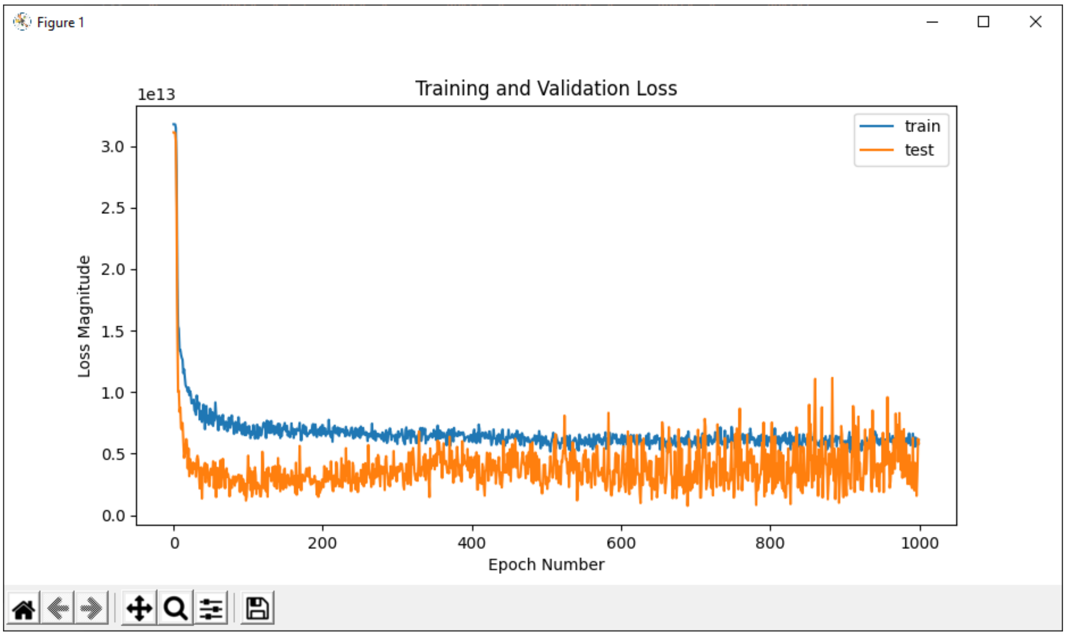
Loss of training and testing data, “total race time” as result label

The error result is concluded by converting the square root of the Mean Squared Error loss to minutes. While the model tried to predict a wide range of total time, the error was unrealistically huge, ranging from 20 minutes to 31 minutes by running models that included various combinations of seasons.



Prediction-answer difference distribution

The result was totally unacceptable in predicting sports like F1, with very thin margin in results. However, we still try to sort all the total race time results from a sample testing race “Abu Dhabi 2019” and see if the final position of the sorted order would resemble the real result of the final position.



Training and testing on 2017-2019 dataset without Abu Dhabi 2019, Test loss: 39.38 minutes

prediction	answer	difference	predicted pos	real final pos
00:56:03.850	01:34:05.715	0.02641048	1	6
00:54:01.740	01:34:22.487	0.02801791	2	2
00:58:21.723	01:34:49.150	0.02531744	3	14
00:58:06.600	01:34:50.940	0.02551319	4	1
00:56:58.466	01:35:10.720	0.02653072	5	7
00:52:24.860	01:35:14.920	0.02974607	6	15
00:56:16.238	01:36:30.618	0.02794421	7	13
00:56:36.300	01:36:31.979	0.02772777	8	8
00:59:15.634	01:36:33.236	0.02589817	9	5
00:57:34.851	01:36:36.836	0.02710631	10	18
00:57:58.691	01:36:37.737	0.02684081	11	10
00:59:04.544	01:36:39.940	0.02610412	12	11
00:56:29.602	01:36:46.495	0.0279733	13	4
00:55:50.536	01:37:01.495	0.02859906	14	3
00:56:18.674	01:37:20.467	0.02849298	15	19
01:01:56.368	01:37:32.482	0.02472354	16	17
00:59:02.200	01:37:43.499	0.02686689	17	12
00:57:02.659	01:40:17.100	0.03002825	18	9
00:58:43.866	01:40:29.473	0.02900008	19	16

Sorted “total time” for predicting “final position”, Abu Dhabi 2019

As shown from the table, we had a total mismatch in final position prediction and the result. The method of using “Total race time” as the result label could be concluded to be ineffective. However, we learned that we could sort a subset of prediction from a race to create a new meaningful prediction.

Version 6: Sorted "final position gained" as the result label

Applying the knowledge of sorting prediction results from a single race, we revisited Version 3 on Abu Dhabi 2019 race. We firstly predicted every racer's final position gain or loss, and then used the initial position of whom to deduct the final position change, which essentially translated to a predicted "final position". Noted that we were not directly predicting "final position" although we had that data column, but we were interpreting the final position information by a simple subtraction of initial position – change in position. After getting each racer's "calculated final position", we sort every result from smallest to largest. The result is as follows:

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
24	16	20	-4	4
0	0	1	1	1
0	0	2	2	2
0	0	3	3	3
0	-1	4	4	5
0	-1	5	5	6
0	-2	6	6	8
0	-4	7	7	11
2	3	10	8	7
0	-2	8	8	10
0	-3	9	9	12
0	-7	11	11	18
2	4	13	11	9
2	0	14	12	14
2	0	15	13	15
2	4	17	15	13
1	0	16	15	16
2	1	18	16	17
2	0	19	17	19

Prediction & answer of final position, Abu Dhabi 2019

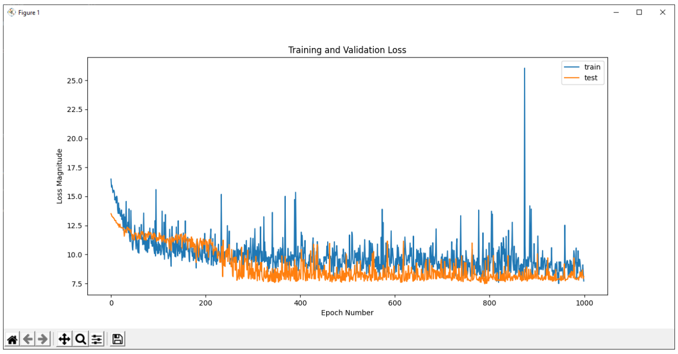
We finally came with a prediction that predicts relatively accurately. However, we have to make further investigations on whether this is an optimal model. After some cross validation checking, we discovered that while the ranked result may seem to be comparatively accurate, the model was still only predicting 0,1,2,3,4 for "final position gain", but not any negative numbers (position loss). The final version must resolve such problem.

Name: Chan Ming Chung
UID: 3035373169

Version 7 (Final version): Sorted "final position gained" with adjusted dataset distribution

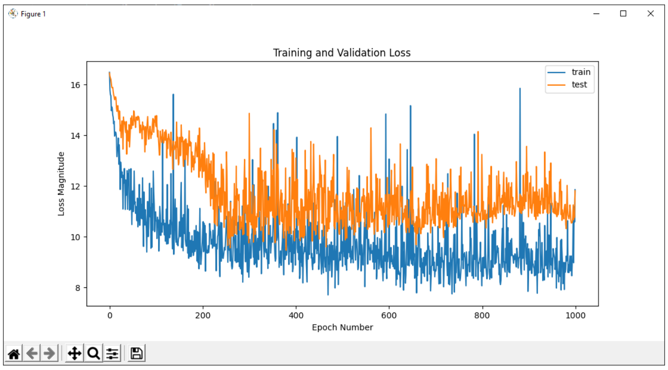
Referencing the strategy in Version 4, we aimed to flatten the distribution of result labels so to encourage the model in predicting diversified numbers. There were 2 strategies adopted in this final version that contributed to the model's success in predicting a wider range of final position changes.

Strategy 1: Optimize the training / testing data split.



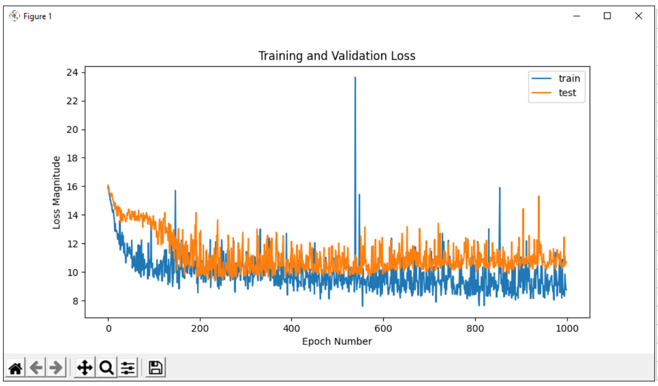
prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
24	16	20	-4	4
0	0	1	1	1
0	0	2	2	2
0	0	3	3	3
0	-1	4	4	5
0	-1	5	5	6
0	-2	6	6	8
0	-4	7	7	11
2	3	10	8	7
0	-2	8	8	10
0	-3	9	9	12
0	-7	11	11	18
2	4	13	11	9
2	0	14	12	14
2	0	15	13	15
2	4	17	15	13
1	0	16	15	16
2	1	18	16	17
2	0	19	17	19

70/30 split of 2016-2019 except the testing target, Abu Dhabi 2019 (Loss: 10.69)



prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
0	0	1	1	1
0	0	2	2	2
0	0	3	3	3
0	-1	4	4	5
0	-1	5	5	6
0	-2	6	6	8
0	-4	7	7	11
0	-2	8	8	10
12	16	20	8	4
1	3	10	9	7
0	-3	9	9	12
0	-7	11	11	18
1	4	13	12	9
1	0	14	13	14
1	0	15	14	15
2	0	16	14	16
2	4	17	15	13
2	1	18	16	17
2	0	19	17	19

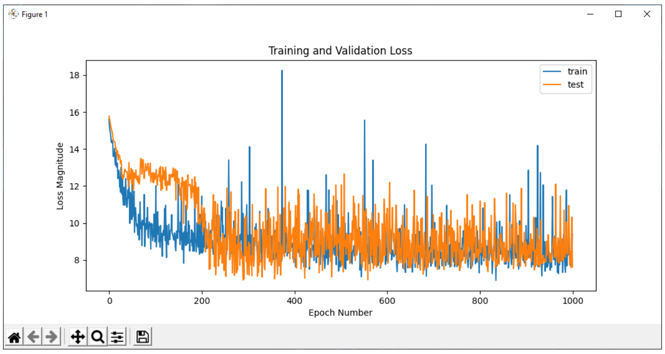
80/20 split of 2016-2019 except the testing target, Abu Dhabi 2019 (Loss: 9.67)



prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
21	16	20	-1	4
0	0	1	1	1
0	0	2	2	2
0	0	3	3	3
0	-1	4	4	5
0	-1	5	5	6
0	-2	6	6	8
0	-4	7	7	11
2	3	10	8	7
0	-2	8	8	10
0	-3	9	9	12
3	4	13	10	9
0	-7	11	11	18
2	0	14	12	14
2	0	15	13	15
2	0	16	14	16
2	4	17	15	13
2	1	18	16	17
2	0	19	17	19

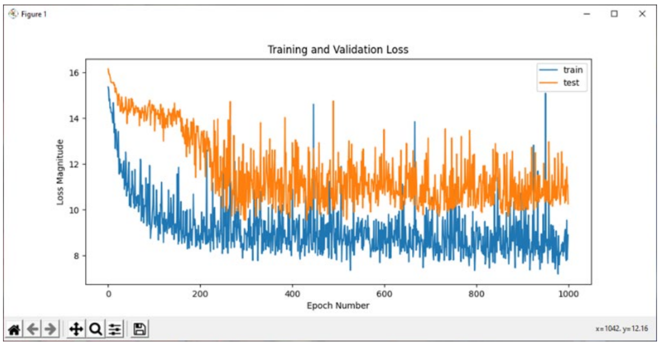
Name: Chan Ming Chung
UID: 3035373169

90/10 split of 2016-2019 except the testing target, Abu Dhabi 2019 (Loss: 10.29)



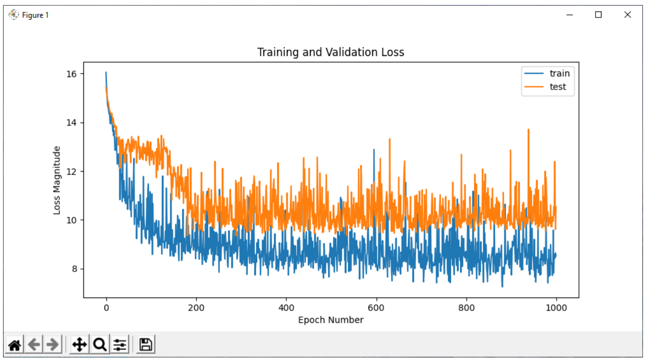
prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
0	0	1	1	1
3	3	5	2	2
0	0	3	3	3
0	0	4	4	4
0	1	6	6	5
0	-1	7	7	8
0	1	8	8	7
0	3	9	9	6
0	-6	10	10	16
1	2	11	10	9
5	6	17	12	11
3	0	15	12	15
0	-6	12	12	18
1	1	13	12	12
7	10	20	13	10
1	1	14	13	13
3	2	16	13	14
4	1	18	14	17

70/30 split of 2016-2019 except the testing target, United States 2019 (Loss: 11.36)



prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
0	0	1	1	1
0	0	3	3	3
2	3	5	3	2
0	0	4	4	4
0	1	6	6	5
1	-1	7	6	8
1	3	9	8	6
0	1	8	8	7
0	-6	10	10	16
1	2	11	10	9
2	1	13	11	12
5	6	17	12	11
3	0	15	12	15
2	1	14	12	13
0	-6	12	12	18
7	10	20	13	10
3	2	16	13	14
4	1	18	14	17

80/20 split of 2016-2019 except the testing target, United States 2019 (Loss: 9.66)

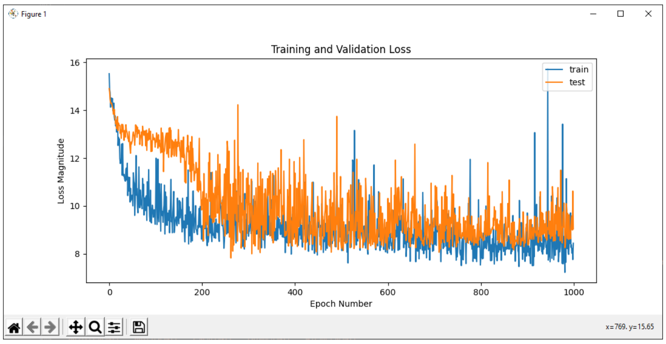


prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
0	0	1	1	1
0	0	3	3	3
0	0	4	4	4
1	3	5	4	2
0	1	6	6	5
0	-1	7	7	8
1	3	9	8	6
0	1	8	8	7
1	-6	10	9	16
0	2	11	11	9
0	-6	12	12	18
1	1	13	12	12
2	0	15	13	15
1	1	14	13	13
3	6	17	14	11
2	2	16	14	14
2	1	18	16	17
3	10	20	17	10

90/10 split of 2016-2019 except the testing target, United States 2019 (Loss: 10.86)

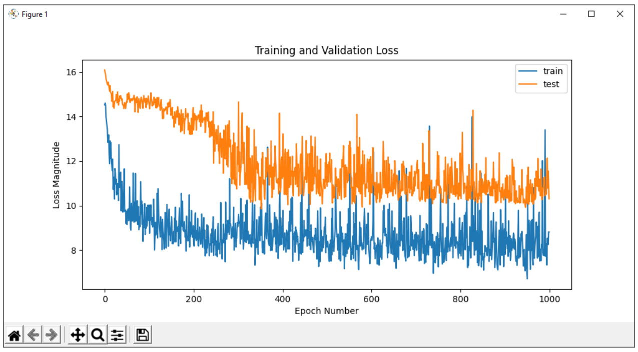
From the results, for both Abu Dhabi and United States, the test loss reached minimum at 80/20 split models. After investigation on other races in 2019 season, we were convinced that unless the test loss worsens to >10, 80/20 split of training and testing dataset would be the optimal splitting ratio.

Name: Chan Ming Chung
UID: 3035373169



prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
0	-3	1	1	4
0	0	2	2	2
0	2	3	3	1
0	-2	4	4	6
2	3	6	4	3
0	0	5	5	5
0	-6	7	7	13
0	-2	9	9	11
0	1	10	10	9
2	5	13	11	8
0	4	11	11	7
0	2	12	12	10
2	4	16	14	12
1	1	15	14	14
2	2	17	15	15
2	1	18	16	17
2	3	19	17	16
1	2	20	19	18

80/20 split of 2016-2019 except the testing target, Mexican 2019 (Loss: 9.34)

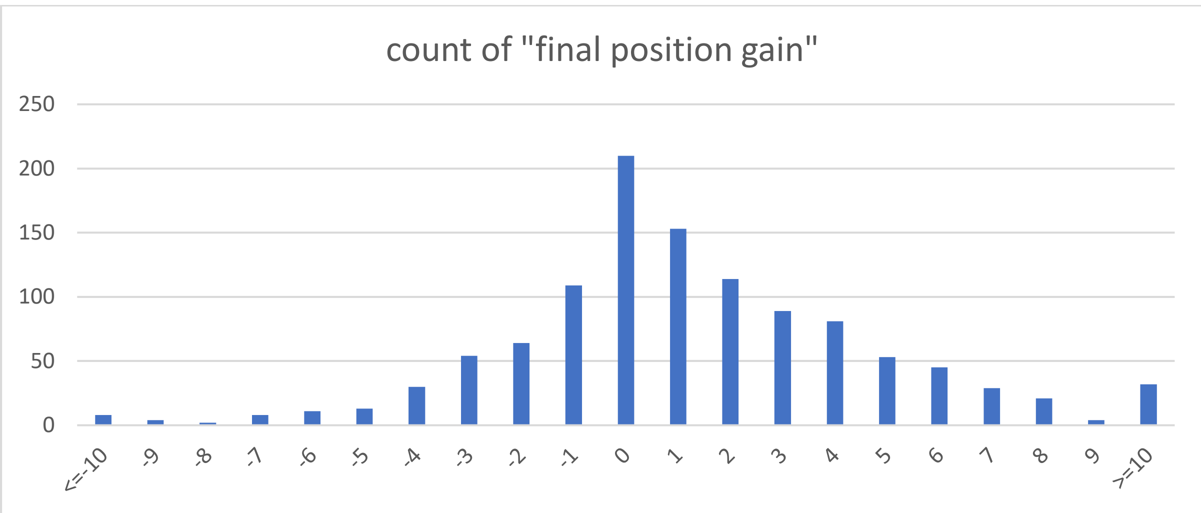


prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
0	-1	1	1	2
0	-4	2	2	6
0	2	3	3	1
0	1	4	4	3
0	2	6	6	4
0	2	7	7	5
0	-3	8	8	11
0	2	9	9	7
0	-3	10	10	13
1	3	12	11	9
0	-3	11	11	14
2	4	14	12	10
0	1	13	13	12
3	2	18	15	16
1	9	17	16	8
2	4	19	17	15
2	3	20	18	17

80/20 split of 2016-2019 except the testing target, Japanese 2019 (Loss: 9.63)

Strategy 2: Remove rows with the highest frequencies at the result label and control the number to be around 50 records per number in “final position gain or loss”

First of all, we visualized the distribution of “final position gain or loss” result labels by the following graph:

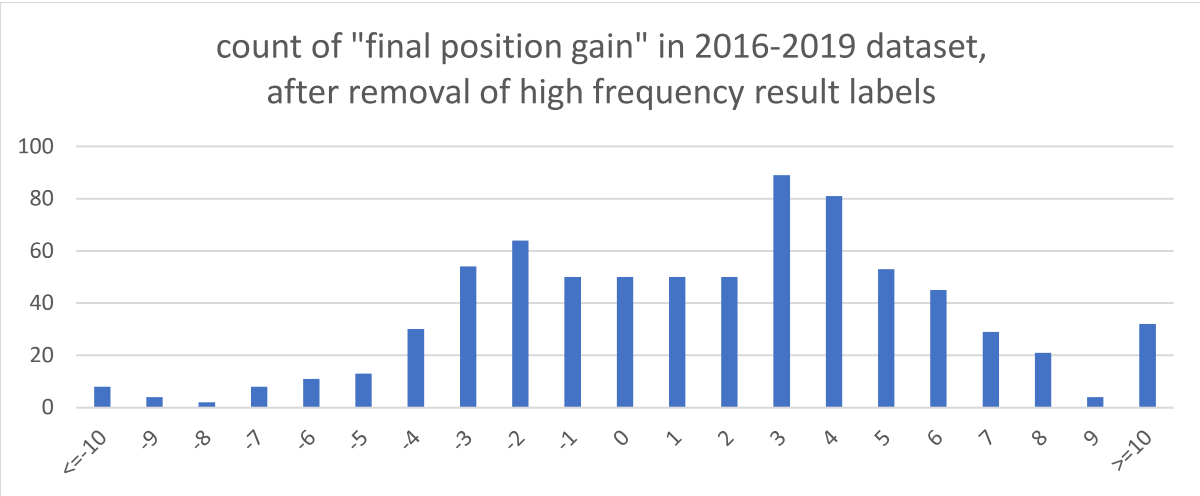


Count of “final position gain” result label distribution in 2016-2019 dataset

Name: Chan Ming Chung
UID: 3035373169

It showed that zero “final position gain” accounted for 210 records, while negative numbers such as -3 (final position loss) had only 54 records, and positive numbers such as 5 (final position gain) had 53 records. Our goal is to flatten the peak near zero such that the distribution could be more balanced.

The methodology applied was to first extract all indices of records with “final position gain” result label as “-1”, “0”, “1”, and “2” separately. Next, we use numpy random.choice function to choose 19, 99, 55, 27 records correspondingly from the list of index for each list. Finally, we dropped the chosen indices for these records. The resulting distribution would be as follows:

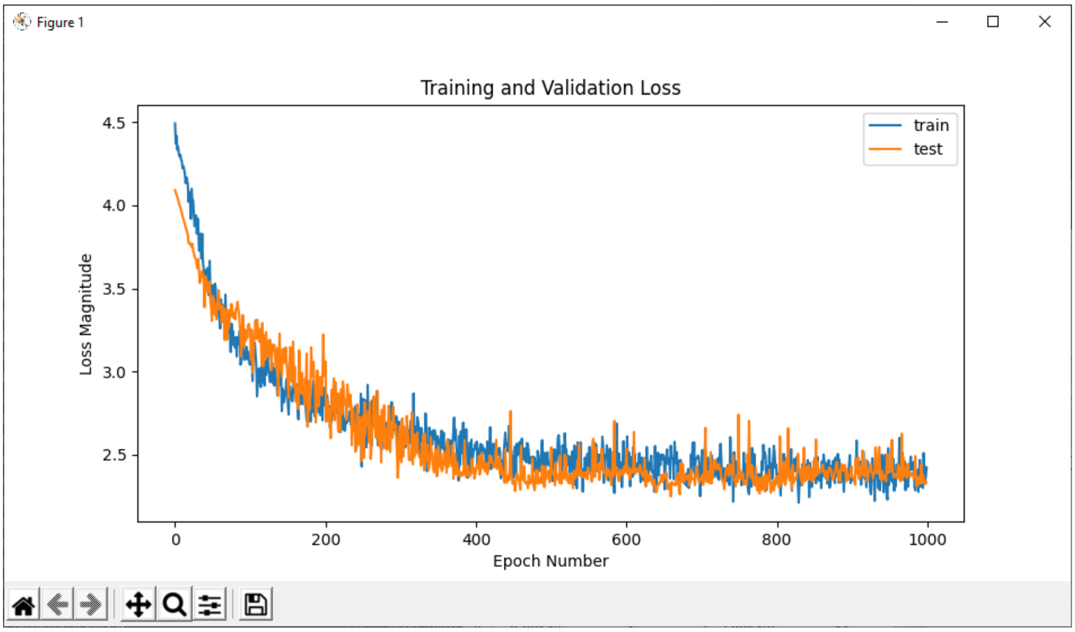


Count of “final position gain” result label distribution in 2016-2019 dataset,

after removing high frequency result labels.

Different extent of index dropping was tested beforehand, and this removal strategy generated excellent and the best performance in terms of predicting various “final position gain or loss” accurately. In addition, we decided to preserve the decimal points of “final position gain or loss” predictions so to eliminate most of the sorting problem where 2 records shared the same “calculated final position”. The results would be as follows:

Name: Chan Ming Chung
UID: 3035373169



Loss of training and testing by using "calculated final position", 2016-2018 seasons.

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.00066	0	1	1.000659	1
-0.00066	0	2	2.000659	2
-0.00066	0	3	3.000659	3
16.48205	16	20	3.517948	4
-0.00066	-1	4	4.000659	5
-0.00066	-1	5	5.000659	6
-0.00066	-2	6	6.000659	8
-0.00066	-4	7	7.000659	11
1.271708	-3	9	7.728292	12
2.117435	3	10	7.882565	7
-0.00066	-2	8	8.000659	10
2.299961	4	13	10.70004	9
-0.00066	-7	11	11.00066	18
1.718063	0	14	12.28194	14
3.160647	0	16	12.83935	16
1.553427	0	15	13.44657	15
2.62107	4	17	14.37893	13
2.583529	1	18	15.41647	17
2.699154	0	19	16.30085	19

"final position" calculated & sorted by predicting "final position gain", Abu Dhabi 2019

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.00066	-1	1	1.000659	2
-0.00066	-4	2	2.000659	6
-0.00066	2	3	3.000659	1
-0.00066	1	4	4.000659	3
-0.00066	2	6	6.000659	4
0.176864	2	7	6.823136	5
1.326783	2	9	7.673217	7
-0.00066	-3	8	8.000659	11
-0.00066	-3	10	10.00066	13
1.964168	3	12	10.03583	9
-0.00066	-3	11	11.00066	14
2.594761	4	14	11.40524	10
0.613811	1	13	12.38619	12
2.165429	9	17	14.83457	8
3.158232	2	18	14.84177	16
5.035952	3	20	14.96405	17
2.846496	4	19	16.1535	15

"final position" calculated & sorted by predicting "final position gain", Japanese 2019

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.00066	0	1	1.000659	1
-0.00066	0	3	3.000659	3
-0.00066	0	4	4.000659	4
0.807034	3	5	4.192966	2
-0.00066	1	6	6.000659	5
2.533227	3	9	6.466773	6
0.475167	-1	7	6.524833	8
-0.00066	1	8	8.000659	7
4.369251	1	14	9.630749	13
-0.00066	-6	10	10.00066	16
-0.00066	2	11	11.00066	9
3.873461	0	15	11.12654	15
5.641126	6	17	11.35887	11
4.43587	2	16	11.56413	14
-0.00066	-6	12	12.00066	18
7.166437	10	20	12.83356	10
-0.00066	1	13	13.00066	12
4.540599	1	18	13.4594	17

"final position" calculated & sorted by predicting "final position gain", United States 2019

The model eventually predicts both positive and negative final position changes, and the calculated "final position" shows a higher accuracy by comparing the custom index, which is calculated by average position difference of all predictions. A cross validation over the 2019 and 2020 season is conducted and further verified that the model is optimized. However, some sporadic race events were too chaotic such that crashing accidents happened too much, causing none of the aforementioned models to predict the race results accurately, such as Brazilian 2019, Belgian 2019, Bahrain 2019, Italian 2020, Tuscan 2020, Bahrain 2020.

Final Model design architecture

For dividing the training data set and testing data set, we decided to split the data set by pandas.sample function to divide 20% of validation data randomly from the entire data source for each epoch during the training process. Afterwards, we used preprocessing.Normalization function from Tensorflow Keras Regularizers module to normalize all feature labels. At each of the 10 layers of the model, we implemented ReLU activation function because ReLU has a linear behavior which makes optimization easier for the neural network. L2 regularizers were also implemented at the rate of 0.001 for each Dense layer, with a Dropout layer at the rate of 0.5 in between each layer. Mean squared error (MSE) was used for the loss function since it is insensitive to outliers and errors, and

RMSprop for optimizer since it is a very robust optimizer and ideal for deepnets. There were **1135 records from 2016-2020** season, and **934 records from 2016-2019 season**. The model architecture would be as follows:

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
normalization (Normalization (None, 17)                35
_________________________________________________________________
dense (Dense)                (None, 256)               4608
_________________________________________________________________
dropout (Dropout)            (None, 256)               0
_________________________________________________________________
dense_1 (Dense)              (None, 256)               65792
_________________________________________________________________
dropout_1 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_2 (Dense)              (None, 128)               32896
_________________________________________________________________
dropout_2 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_3 (Dense)              (None, 128)               16512
_________________________________________________________________
dropout_3 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_4 (Dense)              (None, 64)                8256
_________________________________________________________________
dropout_4 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_5 (Dense)              (None, 64)                4160
_________________________________________________________________
dropout_5 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_6 (Dense)              (None, 32)                2080
_________________________________________________________________
dropout_6 (Dropout)          (None, 32)                0
_________________________________________________________________
dense_7 (Dense)              (None, 16)                528
_________________________________________________________________
dropout_7 (Dropout)          (None, 16)                0
_________________________________________________________________
dense_8 (Dense)              (None, 16)                272
_________________________________________________________________
dropout_8 (Dropout)          (None, 16)                0
_________________________________________________________________
dense_9 (Dense)              (None, 1)                 17
=================================================================
Total params: 135,156
Trainable params: 135,121
Non-trainable params: 35
_________________________________________________________________
```

Model 1: Final model design architecture.

Assessing the accuracy from the race result

The purpose of the first model is to predict the final position of each racer. We isolate the whole 2019 season and train / test the model with only the 2016-2018 seasons for the first test; Next, we isolate the whole 2020 season and train / test the model with only the 2016-2019 seasons. The results would be as follows:

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	0	1	1.000659454	1
-0.000659454	0	2	2.000659454	2
-0.000659454	0	3	3.000659454	3
16.48205185	16	20	3.517948151	4
-0.000659454	-1	4	4.000659454	5
-0.000659454	-1	5	5.000659454	6
-0.000659454	-2	6	6.000659454	8
-0.000659454	-4	7	7.000659454	11
1.271707773	-3	9	7.728292227	12
2.117434502	3	10	7.882565498	7
-0.000659454	-2	8	8.000659454	10
2.299961329	4	13	10.70003867	9
-0.000659454	-7	11	11.00065945	18
1.718063235	0	14	12.28193676	14
3.160647154	0	16	12.83935285	16
1.553426504	0	15	13.4465735	15
2.621070147	4	17	14.37892985	13
2.583529234	1	18	15.41647077	17
2.699153662	0	19	16.30084634	19

Abu Dhabi 2019, accuracy index = 1.368421

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	-1	1	1.000659454	2
-0.000659454	1	2	2.000659454	1
-0.000659454	-1	3	3.000659454	4
-0.000659454	1	4	4.000659454	3
-0.000659454	0	5	5.000659454	5
-0.000659454	1	7	7.000659454	6
-0.000659454	-4	8	8.000659454	12
-0.000646977	1	9	9.000646977	8
-0.000659454	-3	10	10.00065945	13
2.06981039	-1	13	10.93018961	14
-0.000659454	4	11	11.00065945	7
2.279903412	-1	14	11.72009659	15
4.37622118	6	17	12.62377882	11
2.014238358	5	15	12.98576164	10
2.085255861	7	16	13.91474414	9
3.481956959	3	19	15.51804304	16
0.109912314	3	20	19.89008769	17

Australia 2019, accuracy index = 2.117647

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	-1	1	1.000659454	2
-0.000659454	1	2	2.000659454	1
0.872958839	0	3	2.127041161	3
1.625734329	-1	4	2.374265671	5
-0.000659454	-1	5	5.000659454	6
-0.000659454	-3	6	6.000659454	9
1.982416749	1	8	6.017583251	7
-0.000659454	-3	7	7.000659454	10
4.505212784	0	12	7.494787216	12
0.707236111	5	9	8.292763889	4
3.6214149	2	13	9.3785851	11
1.285858393	-5	11	9.714141607	16
-0.000659454	-9	10	10.00065945	19
3.723143339	0	14	10.27685666	14
4.032083511	2	15	10.96791649	13
4.324175835	-1	16	11.67582417	17
6.776958942	11	19	12.22304106	8
5.684755802	3	18	12.3152442	15
4.350969315	-3	17	12.64903069	20
5.381123543	2	20	14.61887646	18

Austrian 2019, accuracy index = 2.3

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	0	1	1.000659454	1
-0.000659454	0	2	2.000659454	2
-0.000659454	0	3	3.000659454	3
-0.000659454	0	4	4.000659454	4
-0.000659454	-1	5	5.000659454	6
-0.000659454	-1	7	7.000659454	8
1.812516689	2	9	7.187483311	7
-0.000659454	3	8	8.000659454	5
2.32362175	0	11	8.67637825	11
2.565809727	4	13	10.43419027	9
0.485318869	-1	12	11.51468113	13
3.016070366	1	15	11.98392963	14
4.323311806	5	17	12.67668819	12
5.998372555	9	19	13.00162745	10
2.587618113	1	16	13.41238189	15
0.000758841	2	18	17.99924116	16

Azerbaijan 2019, accuracy index = 1.125

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	-2	1	1.000659454	3
-0.000659454	-3	2	2.000659454	5
-0.000659454	2	3	3.000659454	1
-0.000659454	2	4	4.000659454	2
-0.000659454	1	5	5.000659454	4
-0.000659454	-7	6	6.000659454	13
-0.000659454	-12	7	7.000659454	19
-0.000659454	1	8	8.000659454	7
3.18825531	3	12	8.81174469	9
1.165140867	-8	10	8.834859133	18
-0.000659454	3	9	9.000659454	6
3.5867033	5	13	9.4132967	8
5.828898907	4	18	12.17110109	14
1.808029771	4	14	12.19197023	10
2.359503984	3	15	12.64049602	12
3.389956474	0	17	13.61004353	17
2.192979097	5	16	13.8070209	11
6.087999344	4	20	13.91200066	16
4.375903606	4	19	14.62409639	15

Bahrain 2019, accuracy index = 3.578947

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
6.756888866	1	3	-3.756888866	2
6.54337883	1	4	-2.54337883	3
2.286286592	0	1	-1.286286592	1
-0.000659454	-2	2	2.000659454	4
3.478235006	1	7	3.521764994	6
3.83181715	-4	8	4.16818285	12
12.70479012	12	17	4.295209885	5
6.074142933	0	11	4.925857067	11
4.051034927	-4	9	4.948965073	13
7.881949425	4	13	5.118050575	9
-0.000659454	-10	6	6.000659454	16
3.868569613	-4	10	6.131430387	14
6.664627552	-1	14	7.335372448	15
3.915492058	4	12	8.084507942	8
10.78610229	12	19	8.213897705	7
9.518606186	0	18	8.481393814	18
6.778029442	6	16	9.221970558	10
10.33220768	3	20	9.66779232	17

Belgian 2019, accuracy index = 3

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	0	1	1.000659454	1
-0.000659454	-4	3	3.000659454	7
-0.000659454	-9	5	5.000659454	14
-0.000659454	4	6	6.000659454	2
-0.000659454	-6	7	7.000659454	13
-0.000659454	4	8	8.000659454	4
-0.000659454	-2	9	9.000659454	11
-0.000659454	2	10	10.00065945	8
0.074556306	5	11	10.92544369	6
4.311087132	6	16	11.68891287	10
-0.000659454	7	12	12.00065945	5
2.527274847	6	15	12.47272515	9
-0.000659454	-2	13	13.00065945	15
4.554410934	17	20	15.44558907	3
0.225014567	6	18	17.77498543	12
-0.000659454	3	19	19.00065945	16

Brazilian 2019, accuracy index = 3.75

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
4.569108963	1	2	-2.569108963	1
-0.000659454	-1	1	1.000659454	2
2.617040157	-1	4	1.382959843	5
-0.000659454	0	3	3.000659454	3
-0.000659454	1	5	5.000659454	4
-0.000659454	-10	6	6.000659454	16
-0.000659454	0	7	7.000659454	7
-0.000659454	-3	8	8.000659454	11
4.149906635	7	13	8.850093365	6
-0.000659454	-3	9	9.000659454	12
7.435114861	8	17	9.564885139	9
-0.000659454	0	10	10.00065945	10
0.326253027	4	12	11.67374697	8
5.227611065	5	19	13.77238894	14
5.756454468	5	20	14.24354553	15
3.395848513	5	18	14.60415149	13
-0.000659454	-2	15	15.00065945	17

British 2019, accuracy index = 2

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	-1	1	1.000659454	2
-0.000659454	1	2	2.000659454	1
-0.000659454	0	3	3.000659454	3
-0.000659454	-2	4	4.000659454	6
-0.000659454	-3	5	5.000659454	8
-0.000659454	2	6	6.000659454	4
-0.000659454	0	7	7.000659454	7
1.752254128	4	9	7.247745872	5
-0.000659454	0	10	10.00065945	10
3.390625477	0	14	10.60937452	14
-0.000659454	0	11	11.00065945	11
0.010875692	-1	12	11.98912431	13
2.983307123	3	15	12.01669288	12
4.479251862	8	17	12.52074814	9
3.598345518	2	18	14.40165448	16
4.241364956	1	19	14.75863504	18
5.036110878	3	20	14.96388912	17
-0.000659454	1	16	16.00065945	15

Canada 2019, accuracy index = 1.66667

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	-1	1	1.000659454	2
-0.000659454	1	2	2.000659454	1
-0.000659454	0	3	3.000659454	3
-0.000659454	-1	4	4.000659454	5
-0.000659454	1	5	5.000659454	4
-0.000659454	0	6	6.000659454	6
-0.000656293	0	7	7.000656293	7
-0.000659454	-4	9	9.000659454	13
-0.000659454	-1	10	10.00065945	11
1.970976114	4	13	11.02902389	9
0.967710972	4	12	11.03228903	8
0.112740546	0	14	13.88725945	14
5.902472496	10	20	14.0975275	10
-0.000659454	-3	15	15.00065945	18
2.741271973	1	18	15.25872803	17
-0.000637839	4	16	16.00063784	12
2.96810627	4	19	16.03189373	15
0.51272732	1	17	16.48727268	16

Chinese 2019, accuracy index = 1.888889

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	0	1	1.000659454	1
-0.000659454	0	2	2.000659454	2
-0.000659454	0	3	3.000659454	3
-0.000659454	0	4	4.000659454	4
-0.000659454	-4	5	5.000659454	9
-0.000659454	0	6	6.000659454	6
-0.000659454	2	7	7.000659454	5
1.83940351	-1	9	7.16059649	10
-0.000659454	-3	8	8.000659454	11
2.145950556	-4	11	8.854049444	15
-0.000659454	-6	10	10.00065945	16
1.740967393	5	12	10.25903261	7
2.550969601	5	13	10.4490304	8
2.325745106	2	14	11.67425489	12
2.323991776	-2	15	12.67600822	17
3.064607143	4	17	13.93539286	13
2.710153103	0	18	15.2898469	18
3.456325531	5	19	15.54367447	14
3.100906849	1	20	16.89909315	19

French 2019, accuracy index = 2.210526

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	-1	1	1.000659454	2
-0.000659454	-6	2	2.000659454	8
-0.000659454	2	3	3.000659454	1
-0.000659454	0	4	4.000659454	4
-0.000659454	2	5	5.000659454	3
-0.000659454	0	6	6.000659454	6
-0.000659454	-2	7	7.000659454	9
0.791037619	3	8	7.208962381	5
2.038635969	2	12	9.961364031	10
-0.000659454	3	10	10.00065945	7
-0.000659454	-1	11	11.00065945	12
1.952772498	-2	13	11.0472275	15
2.103039742	1	14	11.89696026	13
1.9229182	-1	15	13.0770818	16
2.511052132	5	16	13.48894787	11
2.529721737	-1	17	14.47027826	18
4.330314636	6	20	15.66968536	14
2.120003223	1	18	15.87999678	17
2.568767786	0	19	16.43123221	19

Hungarian 2019, accuracy index = 1.894737

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	0	1	1.000659454	1
-0.000659454	-1	2	2.000659454	3
-0.000659454	1	3	3.000659454	2
-0.000659454	-9	4	4.000659454	13
-0.000659454	1	5	5.000659454	4
-0.000659454	1	6	6.000659454	5
12.0463562	11	19	6.953643799	8
-0.000659454	2	8	8.000659454	6
-0.000659454	-3	9	9.000659454	12
-0.000659454	1	10	10.00065945	9
2.375406981	0	14	11.62459302	14
-0.000659454	-3	13	13.00065945	16
2.981223822	6	16	13.01877618	10
3.663789749	6	17	13.33621025	11
3.603523254	11	18	14.39647675	7
0.133658841	-2	15	14.86634116	17
4.045710087	5	20	15.95428991	15

Italian 2019, accuracy index = 2.588235

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	-1	1	1.000659454	2
-0.000659454	-4	2	2.000659454	6
-0.000659454	2	3	3.000659454	1
-0.000659454	1	4	4.000659454	3
-0.000659454	2	6	6.000659454	4
0.176864386	2	7	6.823135614	5
1.32678318	2	9	7.67321682	7
-0.000659454	-3	8	8.000659454	11
-0.000659454	-3	10	10.00065945	13
1.964167595	3	12	10.03583241	9
-0.000659454	-3	11	11.00065945	14
2.594761133	4	14	11.40523887	10
0.613811255	1	13	12.38618875	12
2.165429115	9	17	14.83457088	8
3.158232212	2	18	14.84176779	16
5.035952091	3	20	14.96404791	17
2.846495867	4	19	16.15350413	15

Japanese 2019, accuracy index = 2

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	-3	1	1.000659454	4
-0.000659454	0	2	2.000659454	2
-0.000659454	2	3	3.000659454	1
-0.000659454	-2	4	4.000659454	6
1.26859796	3	6	4.73140204	3
-0.000659454	0	5	5.000659454	5
-0.000659454	-6	7	7.000659454	13
-0.000659454	-2	9	9.000659454	11
-0.000659454	1	10	10.00065945	9
0.881905675	4	11	10.11809433	7
2.859827757	5	13	10.14017224	8
1.086053967	2	12	10.91394603	10
2.296399832	1	15	12.70360017	14
2.822880268	4	16	13.17711973	12
2.588597775	2	17	14.41140223	15
2.938324451	1	18	15.06167555	17
2.56802702	3	19	16.43197298	16
2.132705212	2	20	17.86729479	18

Mexican 2019, accuracy index = 1.777778

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
7.74551487	-1	2	-5.74551487	3
-0.000659454	0	1	1.000659454	1
-0.000641903	-1	3	3.000641903	4
-0.000659454	2	4	4.000659454	2
-0.000659454	-9	5	5.000659454	14
-0.000659454	-3	6	6.000659454	9
-0.000659454	0	7	7.000659454	7
-0.000659454	3	8	8.000659454	5
1.933865309	2	10	8.066134691	8
-0.000650481	3	9	9.000650481	6
2.776863813	1	12	9.223136187	11
3.021641254	3	13	9.978358746	10
3.137068987	-3	14	10.86293101	17
-0.000659454	-2	11	11.00065945	13
2.467739344	4	16	13.53226066	12
3.381489277	1	17	13.61851072	16
3.760550499	-1	18	14.2394495	19
2.528172016	4	19	16.47182798	15
3.340728283	2	20	16.65927172	18

Monaco 2019, accuracy index = 2.210526

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	-2	1	1.000659454	3
-0.000659454	1	2	2.000659454	1
1.108295441	2	4	2.891704559	2
-0.000659454	-1	5	5.000659454	6
3.716090679	5	9	5.283909321	4
-0.000659454	-4	6	6.000659454	10
-0.000659454	-1	7	7.000659454	8
1.89935112	4	11	9.10064888	7
2.264422417	4	13	10.73557758	9
2.213954449	3	14	11.78604555	11
-0.000659454	-3	12	12.00065945	15
6.77711916	7	19	12.22288084	12
7.686756134	15	20	12.31324387	5
2.247176886	2	15	12.75282311	13
2.87524581	2	16	13.12475419	14

Russian 2019, accuracy index = 1.866667

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	-1	1	1.000659454	2
-0.000659454	-2	2	2.000659454	4
-0.000659454	2	3	3.000659454	1
-0.000659454	1	4	4.000659454	3
-0.000659454	0	5	5.000659454	5
-0.000659454	0	6	6.000659454	6
-0.000659454	-5	7	7.000659454	12
-0.000659454	-1	8	8.000659454	9
2.467889786	3	11	8.532110214	8
-0.000659454	2	9	9.000659454	7
-0.000655855	0	10	10.00065586	10
-0.000659454	-4	13	13.00065945	17
3.756094694	6	17	13.24390531	11
-0.000659454	-1	14	14.00065945	15
3.48031497	3	19	15.51968503	16
-0.000659454	3	16	16.00065945	13
3.795790911	6	20	16.20420909	14

Singapore 2019, accuracy index = 1.882353

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	-1	1	1.000659454	2
-0.000659454	1	2	2.000659454	1
-0.000659454	-1	3	3.000659454	4
-0.000659454	1	4	4.000659454	3
-0.000659454	0	5	5.000659454	5
-0.000659454	0	6	6.000659454	6
-0.000659454	-3	7	7.000659454	10
-0.000659454	1	8	8.000659454	7
-0.000659454	0	9	9.000659454	9
1.558761835	4	12	10.44123816	8
-0.000659454	0	11	11.00065945	11
0.889951527	1	13	12.11004847	12
3.684277773	-1	17	13.31572223	18
-0.000659454	0	14	14.00065945	14
0.372632682	0	15	14.62736732	15
4.18924427	2	19	14.81075573	17
2.984582901	2	18	15.0154171	16
3.192691088	7	20	16.80730891	13

Spanish 2019, accuracy index = 1.222222

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.000659454	0	1	1.000659454	1
-0.000659454	0	3	3.000659454	3
-0.000659454	0	4	4.000659454	4
0.807034194	3	5	4.192965806	2
-0.000659454	1	6	6.000659454	5
2.533227444	3	9	6.466772556	6
0.475166589	-1	7	6.524833411	8
-0.000659454	1	8	8.000659454	7
4.369251251	1	14	9.630748749	13
-0.000659454	-6	10	10.00065945	16
-0.000659454	2	11	11.00065945	9
3.873460531	0	15	11.12653947	15
5.641126156	6	17	11.35887384	11
4.435869694	2	16	11.56413031	14
-0.000659454	-6	12	12.00065945	18
7.166437149	10	20	12.83356285	10
-0.000659454	1	13	13.00065945	12
4.540598869	1	18	13.45940113	17

United States 2019, accuracy index = 2.111111

In the 2019 season, the total average accuracy index is 2.127986, which in other words the final position prediction would have ±2.127986 error.

For the 2020 season, we adjusted the number of highest frequency rows to be randomly removed, but the principle of cutting the peak (at -1 to 2) to 50 records per result label remained the same.

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216752723	-2	1	1.216752723	3
-0.216750711	0	2	2.216750711	2
-0.216749102	3	4	4.216749102	1
-0.216750026	0	6	6.216750026	6
-0.216750249	-4	7	7.216750249	11
-0.216293469	4	8	8.216293469	4
-0.216745719	4	9	9.216745719	5
2.119436264	-1	12	9.880563736	13
-0.198782146	1	10	10.19878215	9
-0.216726571	-1	11	11.21672657	12
2.672385931	6	14	11.32761407	8
-0.216749221	-3	13	13.21674922	16
2.367998838	6	16	13.63200116	10
3.451060534	5	20	16.54893947	15
1.230385184	2	19	17.76961482	17

70th anniversary 2020, accuracy index = 2.4

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216749996	0	1	1.216749996	1
-0.216750011	0	2	2.216750011	2
-0.216740891	0	3	3.216740891	3
-0.216745123	-1	4	4.216745123	5
-0.216714308	1	5	5.216714308	4
0.284086227	0	6	5.715913773	6
1.866585374	-2	8	6.133414626	10
2.443427324	1	9	6.556572676	8
-0.216740951	-4	7	7.216740951	11
1.902515054	1	10	8.097484946	9
1.321264863	4	11	9.678735137	7
2.031000614	-1	12	9.968999386	13
2.295673132	-1	13	10.70432687	14
2.143335581	-2	14	11.85666442	16
2.614485979	3	15	12.38551402	12
2.858964205	1	16	13.1410358	15
2.735896349	1	18	15.26410365	17
2.677708626	2	20	17.32229137	18

Abu Dhabi 2020, accuracy index = 1.111111

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216752961	0	1	1.216752961	1
3.4162395	1	5	1.5837605	4
-0.216753468	0	3	3.216753468	3
-0.216753408	-9	4	4.216753408	13
-0.216751069	0	6	6.216751069	6
-0.216751173	5	7	7.216751173	2
-0.216749594	3	8	8.216749594	5
3.677821636	5	12	8.322178364	7
4.03810358	1	13	8.96189642	12
4.870893002	6	14	9.129106998	8
-0.205650315	1	11	11.20565031	10
6.393766403	9	20	13.6062336	11
3.258857012	9	18	14.74114299	9

Austrian 2020, accuracy index = 2.307692

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216748968	0	1	1.216748968	1
-0.216751635	-6	2	2.216751635	8
-0.216750935	1	4	4.216750935	3
-0.216751322	-13	5	5.216751322	18
-0.216752946	-1	6	6.216752946	7
-0.21675019	-2	7	7.21675019	9
-0.216745406	2	8	8.216745406	6
0.520559847	5	9	8.479440153	4
1.645713687	-2	11	9.354286313	13
-0.216749325	-1	10	10.21674933	11
1.663342714	2	12	10.33665729	10
2.781270504	10	15	12.2187295	5
-0.216746345	2	14	14.21674635	12
-0.216748983	0	16	16.21674898	16
0.206911132	2	17	16.79308887	15
2.832933664	6	20	17.16706634	14

Bahrain 2020, accuracy index = 3

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216752619	0	1	1.216752619	1
-0.21675308	0	2	2.21675308	2
-0.216752544	0	3	3.216752544	3
-0.216752529	0	4	4.216752529	4
-0.216751739	-1	5	5.216751739	6
-0.216752842	1	6	6.216752842	5
-0.216748416	-2	8	8.216748416	10
3.361842871	4	12	8.638157129	8
-0.216748342	0	9	9.216748342	9
2.918996811	-1	13	10.08100319	14
-0.204443514	3	10	10.20444351	7
-0.216745093	0	11	11.21674509	11
2.765480042	1	14	11.23451996	13
3.283275127	4	16	12.71672487	12
3.765510082	2	17	13.23448992	15
4.048149586	3	19	14.95185041	16
4.910264015	3	20	15.08973598	17

Belgian 2020, accuracy index = 0.941176

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216751933	0	1	1.216751933	1
-0.216752514	-9	2	2.216752514	11
-0.216751412	1	3	3.216751412	2
-0.216749623	1	4	4.216749623	3
-0.216747597	0	5	5.216747597	5
2.628846645	4	8	5.371153355	4
-0.216745749	-3	6	6.216745749	9
0.24621968	-6	7	6.75378032	13
2.722108603	0	10	7.277891397	10
2.715027809	4	11	8.284972191	7
0.184799716	3	9	8.815200284	6
0.527024686	4	12	11.47297531	8
3.40498805	1	15	11.59501195	14
4.346184254	1	17	12.65381575	16
2.592652559	-1	16	13.40734744	17
4.354539394	3	18	13.64546061	15
5.2586689	8	20	14.7413311	12

British 2020, accuracy index = 2.588235

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216749191	1	2	2.216749191	1
-0.216749743	1	3	3.216749743	2
-0.216752693	-3	4	4.216752693	7
-0.216747701	3	6	6.216747701	3
1.885223269	5	9	7.114776731	4
2.0162673	5	10	7.9837327	5
3.352033854	6	12	8.647966146	6
2.464660168	-2	13	10.53533983	15
0.324550569	0	11	10.67544943	11
4.309258461	7	16	11.69074154	9
1.353043079	2	15	13.64695692	13
0.229797676	4	14	13.77020232	10
5.825821877	12	20	14.17417812	8
3.502489567	4	18	14.49751043	14
4.495948792	7	19	14.50405121	12

Eifel 2020, accuracy index = 2

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216753468	-1	1	1.216753468	2
-0.216751143	1	2	2.216751143	1
-0.216751844	2	5	5.216751844	3
-0.216752648	-9	6	6.216752648	15
-0.216749981	2	7	7.216749981	5
-0.216751814	4	8	8.216751814	4
-0.216746747	1	9	9.216746747	8
-0.216742143	3	10	10.21674214	7
3.001164198	2	14	10.9988358	12
-0.027931556	5	11	11.02793156	6
3.062059641	9	18	14.93794036	9
4.700881958	10	20	15.29911804	10
3.597006083	8	19	15.40299392	11
-0.197146162	2	16	16.19714616	14

Imola 2020, accuracy index = 2.142857

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216752782	-6	1	1.216752782	7
-0.216750175	-3	2	2.216750175	5
-0.216751859	1	3	3.216751859	2
-0.216750026	-6	4	4.216750026	10
3.929543734	-6	9	5.070456266	15
4.838946342	2	11	6.161053658	9
-0.216746345	2	6	6.216746345	4
0.749712825	1	7	6.250287175	6
3.937840223	4	12	8.062159777	8
-0.216737866	5	8	8.216737866	3
1.627987027	9	10	8.372012973	1
5.184542179	4	16	10.81545782	12
2.634177685	1	14	11.36582232	13
6.143765926	5	19	12.85623407	14
4.354349136	2	18	13.64565086	16
4.457866192	9	20	15.54213381	11

Italian 2020, accuracy index = 3.625

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216748625	0	1	1.216748625	1
-0.216749892	0	2	2.216749892	2
-0.216750517	0	3	3.216750517	3
-0.21675095	0	4	4.21675095	4
-0.216753468	-2	5	5.216753468	7
-0.21675247	-6	6	6.21675247	12
-0.211636454	1	7	7.211636454	6
1.702151418	4	9	7.297848582	5
2.489715338	1	10	7.510284662	9
-0.216751739	-5	8	8.216751739	13
1.739137888	3	11	9.260862112	8
3.268637419	0	14	10.73136258	14
3.83997941	5	15	11.16002059	10
3.771962404	5	16	12.2280376	11
-0.216668889	-6	13	13.21666889	19
3.666806698	2	17	13.3331933	15
3.869149923	1	18	14.13085008	17
3.843741655	3	19	15.15625834	16
4.251885414	2	20	15.74811459	18

Portuguese 2020, accuracy index = 1.789474

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216746897	-2	1	1.216746897	3
-0.216747165	0	2	2.216747165	2
-0.216666669	2	3	3.216666669	1
-0.216743976	0	4	4.216743976	4
-0.216745123	0	5	5.216745123	5
-0.216748908	0	7	7.216748908	7
2.040391207	4	10	7.959608793	6
-0.216749892	-7	8	8.216749892	15
-0.17565988	0	9	9.17565988	9
4.770338535	5	15	10.22966146	10
0.70122534	3	11	10.29877466	8
2.821401834	1	14	11.17859817	13
2.777936697	-1	16	13.2220633	17
3.085854292	6	18	14.91414571	12
1.824532628	6	17	15.17546737	11
3.607550383	5	19	15.39244962	14
3.869446516	4	20	16.13055348	16

Russian 2020, accuracy index = 1.764706

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.132227063	-7	1	1.132227063	8
4.701872349	2	7	2.298127651	5
4.403853416	4	8	3.596146584	4
4.794525146	-2	9	4.205474854	11
-0.216751799	4	5	5.216751799	1
4.387393475	7	10	5.612606525	3
6.196333408	6	12	5.803666592	6
4.88839817	9	11	6.11160183	2
-0.21674937	-1	6	6.21674937	7
5.155132771	1	14	8.844867229	13
4.299708843	0	15	10.70029116	15
7.619794846	9	19	11.38020515	10

Sakhir 2020, accuracy index = 3.583333

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216750443	0	1	1.216750443	1
-0.21675168	1	3	3.21675168	2
-0.216749981	-1	4	4.216749981	5
2.904507875	-2	8	5.095492125	10
-0.216752484	1	5	5.216752484	4
1.00209403	1	7	5.99790597	6
-0.216752559	-2	6	6.216752559	8
2.621538401	1	10	7.378461599	9
2.66195941	4	11	8.33804059	7
2.782933474	0	12	9.217066526	12
2.726452351	2	13	10.27354765	11
3.026889086	2	15	11.97311091	13
2.396545887	1	16	13.60345411	15
3.913892031	1	18	14.08610797	17
-0.189638793	0	14	14.18963879	14
2.518524408	-2	17	14.48147559	19
2.688415289	1	19	16.31158471	18
3.29788661	4	20	16.70211339	16

Spanish 2020, accuracy index = 1.555556

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216752186	0	1	1.216752186	1
-0.216753468	-1	2	2.216753468	3
-0.216753468	-6	3	3.216753468	9
-0.216747046	2	4	4.216747046	2
-0.216751218	2	6	6.216751218	4
-0.216753468	-8	7	7.216753468	15
-0.216746777	0	8	8.216746777	8
-0.216547415	4	9	9.216547415	5
2.667203188	5	12	9.332796812	7
-0.216750473	-5	11	11.21675047	16
1.738682985	3	13	11.26131701	10
4.816385746	11	17	12.18361425	6
3.511567116	5	16	12.48843288	11
1.721457601	3	15	13.2785424	12
3.958894968	1	18	14.04110503	17
4.07671833	5	19	14.92328167	14
4.24495554	7	20	15.75504446	13

Styrian 2020, accuracy index = 2.941176

prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
-0.216747075	0	1	1.216747075	1
-0.216752782	0	2	2.216752782	2
-0.216753125	1	4	4.216753125	3
-0.216753468	-3	5	5.216753468	8
-0.216752127	2	7	7.216752127	5
-0.216751322	4	8	8.216751322	4
4.889224529	3	15	10.11077547	12
0.809102535	5	12	11.19089746	7
-0.216647878	5	11	11.21664788	6
5.669912338	7	18	12.33008766	11
-0.216752157	4	13	13.21675216	9
-0.216750786	4	14	14.21675079	10

Tuscan 2020, accuracy index = 1.666667

In the 2020 season, the total average accuracy index is 2.227799, which in other words the final position prediction would have ±2.227799 error.

An important fact about predicting the 2020 season is that Formula One has introduced a brand new "Corona calendar" with 5 new tracks that nobody has ever raced before. It demonstrates that how our generalization of tracks by their respective characteristics effectively helps to predict not only traditional race circuits, but even new racetracks as well. This is an important indication of how accurate, reliable, and useful this model is to the Scuderia Ferrari Team for extensive future use.

Machine Learning Model 2: Predict pitstop strategy under Safety Car condition

Similar to Model 1, we adopted Tensorflow Keras Sequential model for predicting the final position gain or loss under two scenarios: the racer goes in for a pitstop when the safety car is deployed; Or the racer stays out on the track, together with some race facts at the moment when the safety car is deployed. This is a supervised regression type machine learning model where we feed the model with 15 input factors and 1 result label about the gain or loss in the final position. Before proceeding to the model design evolution, more explanation and discussion of the "safety car scenario" would be provided first.

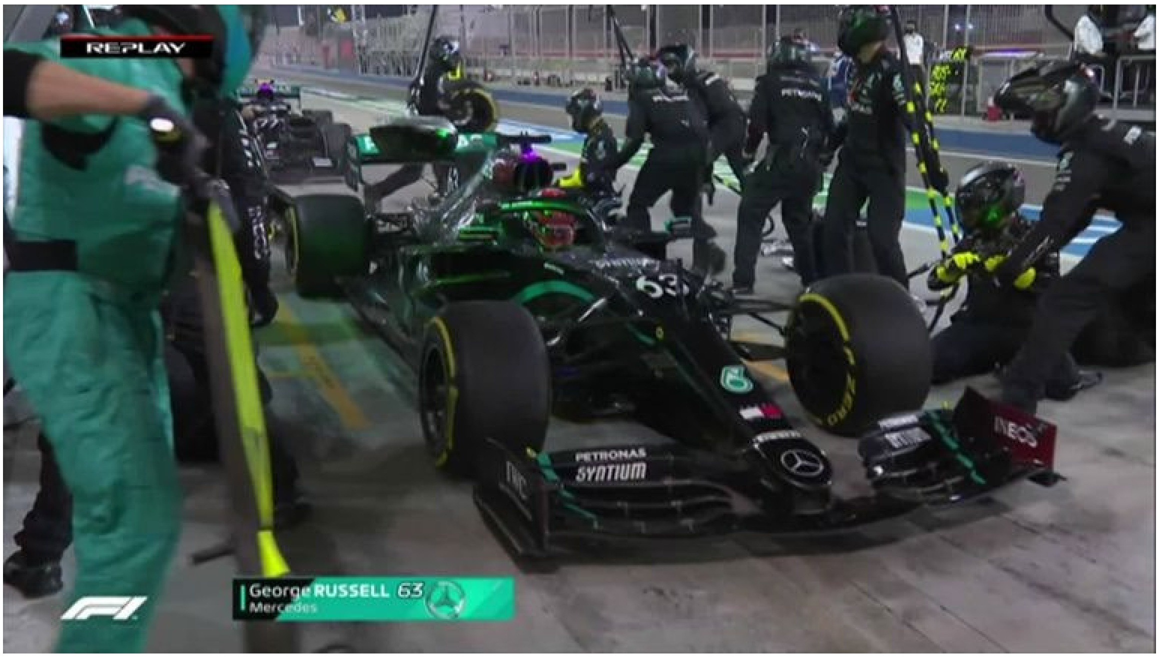
Pitstop strategy difference under Accident Scenario

Formula One is never a peaceful sport. Especially since the race adopted a standing start, where all cars start from a standing grid and accelerate at the same time, there are always side-to-side battles and car crashing during the opening lap, or even the first corner of the track (a.k.a. first lap drama). Whenever there is a hazard caused by accidents on the track, such as a sudden engine failure of a car, car crashing to a wall or with other cars, a yellow flag will be raised, and either a safety car will be deployed to lead the race in a relatively slower pace, or a Virtual Safety Car (VSC) will be deployed where all drivers are forced to drive slowly by maintaining a delta, which is a speed restrictor that limits racing pace. In either of the situations, all cars are required to drive at a slower speed to let the marshal (staff members of the track) clear off the debris and all other blockages on the track.

A safety car leading all drivers to drive under a safe speed limit.

Under such a scenario, a lot of drivers would decide to alter their original planned pitstop strategy and pit in earlier than planned. The reason is that during a normal racing lap, the time spent on going into the pit stop and change tyres could be around 30 seconds, and after the pit stop, the driver would be demoted to a position 30 seconds behind his current position, which would always be an immense drop in positions since the time difference between each driver is very small, accounting for few milliseconds difference per lap; However, under the Accident scenario where the safety car or VSC is deployed, all cars running on the track must slow down, such that after going into a pit stop, the driver will **<u>not</u>** be demoted 30 seconds behind, but most likely halved or even less, because the rivals who are originally 30 seconds behind the driver must drive slowly and therefore cannot catch up the driver going into the pit stop, making that 30 seconds cost virtually less time-consuming.

However, it remains controversial whether the driver should change its planned pitstop strategy and get a “free pit-stop”, because there is still position loss to suffer from going into a pitstop. Some circuits are harder to make overtaking maneuvers and therefore holding position is more important than having fresher tyres for higher car performance. It creates a dilemma for the team and the drivers whether they should sacrifice track position for fresher tyres and better grip, hence better car performance, or stay out at the track to possibly gain some positions due to others choosing to go into the pitstop. The model is therefore aimed to evaluate the benefits of making a pitstop under such an Accident Scenario quickly as to let the team reacts as soon as the safety car is deployed, hence maximize the pit crew members’ time for preparing the tyres and get ready in position.



With little time to react, Mercedes Benz pit crew made a serious mistake by mixing up wrong tyre compounds for their 2 drivers, Sakhir GP 2020.

Dataset description

The list of input factors are as follows:

1. Initial starting position of the racer.
2. Lap number when the safety car is deployed.

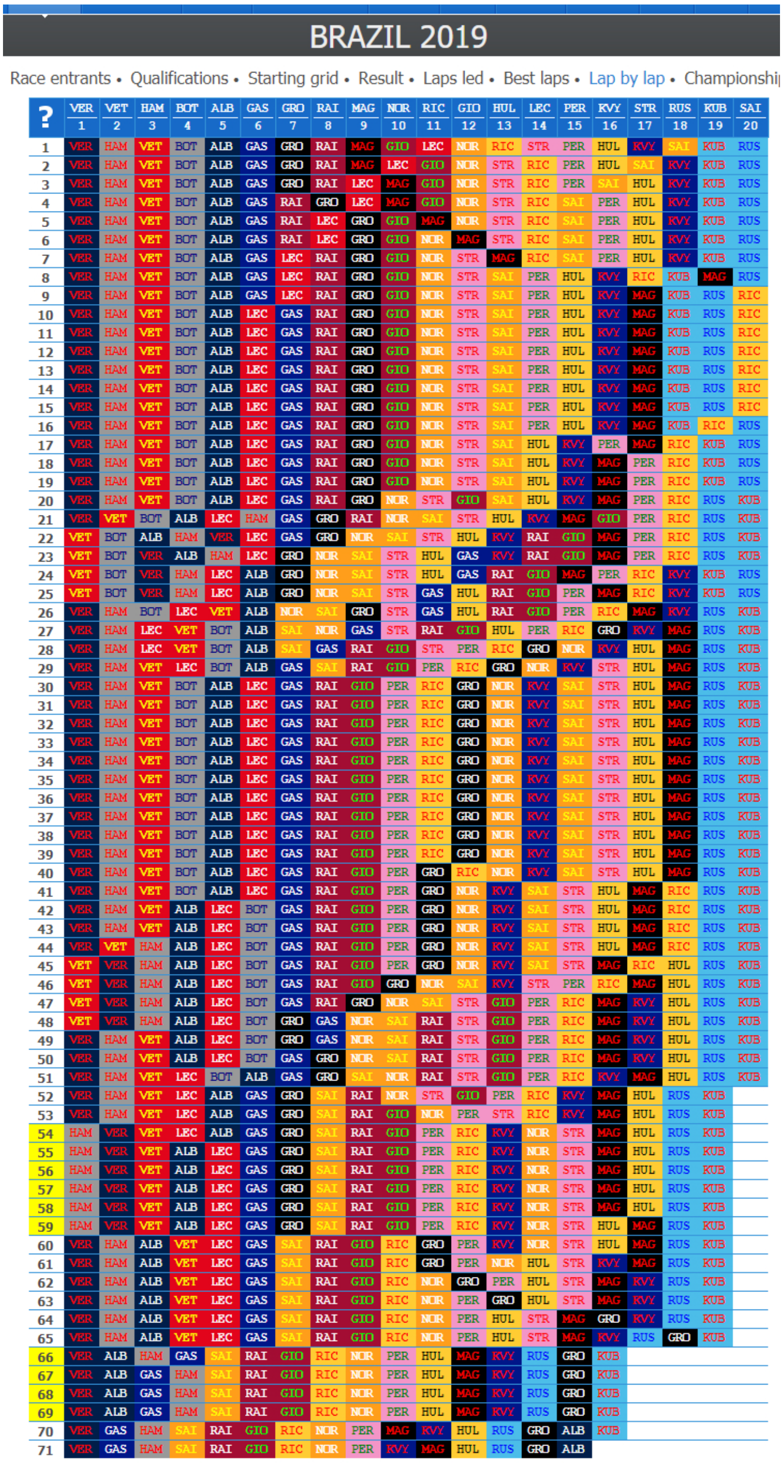
3. Laps travelled for the current tyre set.
4. Laps remained of the race.
5. Current tyre compound.
6. Current position
7. Track temperature.
8. Track humidity.
9. Track maximum altitude change.
10. Track number of turns in a lap.
11. Track race distance / total number of laps.
12. Track total length of race (in km).
13. Team Ability index (from previous season constructors' championship points).
14. Driver Ability index (from previous season drivers' championship points).
15. Binary decision of whether the car goes in for a pitstop.

Our training label (result) is also the final position gain or loss, which reflects on how good different strategies are. By varying the input column of "binary decision of whether the car goes in for a pitstop", we compare the two predicted final position gain as to understand which decision would benefit the racer the most, whether to go into the pitstop or not when the safety car is deployed.

Preprocessing of the dataset

While most of the data columns adopted in this model are derived from the previous tables gathered online, we have to make several changes to the dataset to suit the use of our model.

Name: Chan Ming Chung
UID: 3035373169



Brazil 2019, lap by lap chart showing 2 safety car instances.

For instance, in the race Brazil 2019, there are 2 separate occasions where the safety car is deployed, on lap 54 and lap 66 respectively (cells highlighted in yellow). Therefore, the challenge is to represent a race with one or more than one safety car deployment, while maintaining a constant input shape of the machine learning model. Therefore, we consider the two or more safety car instances as separate events and treat each safety car deployment as a single row of record. In Brazil 2019, we duplicated every racer's record twice so to differentiate the pitstop strategy of the one on lap 54 and the one on lap 66.

No.	initial_pos	sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	final_pos_gained
3	11	54	0	14	17	3	12	5
4	10	54	0	1	17	2	14	2
7	8	54	0	7	17	3	9	4
8	7	54	0	28	17	2	7	-6
10	6	54	0	7	17	3	6	4
11	15	54	0	9	17	3	11	6
20	9	54	0	1	17	3	16	-2
23	5	54	0	4	17	3	5	-9
26	16	54	0	8	17	3	13	6
27	13	54	0	1	17	3	17	-2
33	1	54	1	10	17	2	2	0
44	3	54	0	11	17	3	1	-4
55	20	54	0	25	17	2	8	17
63	18	54	1	31	17	2	18	6
88	19	54	0	1	17	1	19	3
99	12	54	0	8	17	3	10	7
3	11	66	0	26	5	3	8	5
4	10	66	0	13	5	2	9	2
7	8	66	0	19	5	3	6	4
8	7	66	0	1	5	3	15	-6
10	6	66	0	19	5	3	4	4
11	15	66	0	21	5	3	10	6
20	9	66	0	13	5	3	12	-2
23	5	66	0	16	5	3	2	-9
26	16	66	0	20	5	3	13	6
27	13	66	0	13	5	3	11	-2
33	1	66	0	12	5	2	1	0
44	3	66	1	23	5	2	3	-4
55	20	66	0	37	5	2	5	17
63	18	66	1	12	5	3	14	6
88	19	66	1	13	5	3	16	3
99	12	66	0	20	5	3	7	7

Brazil 2019, showing duplicated racers' data rows.

For example, racer number 33 decided to go into the pitstop during the first safety car instance, but not the second time; while racer number 44 stayed out for the first safety car instance but went in for a pitstop at the second time of the race. In this fashion of displaying each data record, we focus on **the observable race facts when the safety car is deployed**, rather than the whole race results, which is logically yet to happen. In addition, a pitstop made previously (e.g. on lap 54) does not matter to the current pitstop strategy decision (e.g. on lap 66), since the model understands the current state and condition of the racecar, rather than the historical pitstop strategies that lead to the current state.

Model design evolutions

Version 1: Reusing model 1 dataset with slight changes.

Taking advantage of the first model that predicts the final position, we made slight changes to the model to predict the optimal pitstop strategy under the accident scenario.

No.	initial_pos	final_pos	no_of_pits	tyre_grid	tyre_1	tyre_2	tyre_grid_distance	tyre_1_distance	tyre_2_distance	sc (decision made)	final_pos_gained
3	6	9		5			11			1	-3
4	12	11		4			11			0	1
5	4	2		5			11			1	2
7	14	17		5			11			0	-3
8	13	10		5			11			0	3
10	8	5		5			11			0	3
11	16	12		4			11			1	4
16	15	DNF		4			11			0	
18	17	16		4			11			0	1
20	5	14		5			11			1	-9
23	10	8		5			11			0	2
26	7	7		5			11			0	0
27	11	13		4			11			0	-2
33	3	4		5			11			1	-1
44	1	1		5			11			1	0
55	9	6		5			11			0	3
63	19	15		5			11			0	4
77	2	3		5			11			2	-1
88	20	18		4			11			0	2
99	18	19		5			11			0	-1

Dataset right after the first accident by No. 16 Charles Leclerc, Monaco 2019.

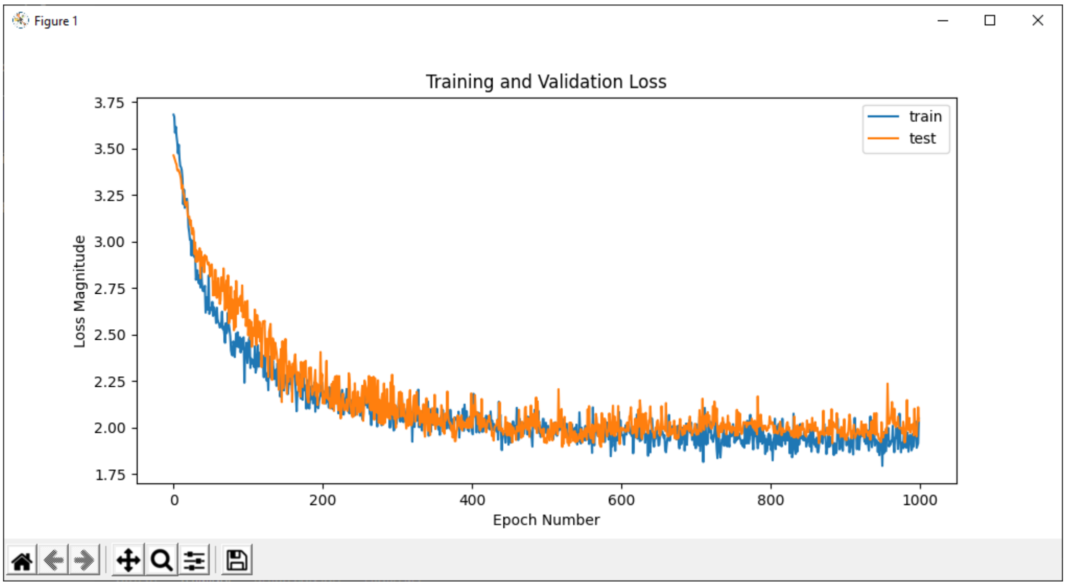
As shown from the table, we will have the starting tyre compound, and the number of laps traveled on this set of tyres. The column named “sc” means whether the driver decided to go into the pit stop for a new set of tyres, with >=1 meaning yes, and 0 meaning no. Treating these as the input data, together with weather and track information, driver and team ability index, we could predict the “final position gain” label. This model would imply that the pitstop strategist could just input 0 and 1 to the “sc” column and run this model twice to get 2 results on the predicted final position gain and choose the one with a better-predicted result.

However, for races that have more than one safety car instance, this model cannot represent both safety car cases. In addition, the column "sc" summarized the total number of pitstop made under the safety car condition, which means if there are more than one occasion of safety car deployment, the number 1 could be either from the first safety car instance, or the second one. The "sc" column is therefore too ambiguous for our usage in this model.

Version 2 (Final version): Adding derived data columns and duplicating rows of racer records.

As mentioned from the previous "preprocessing of dataset" section, we added the decision of pit in or stay out, laps travelled for current tyre set, laps remained, current tyre compound and position before safety car condition as new input data columns of the dataset. We also duplicated rows for races with multiple safety car conditions to show pitstop strategies at different instances.

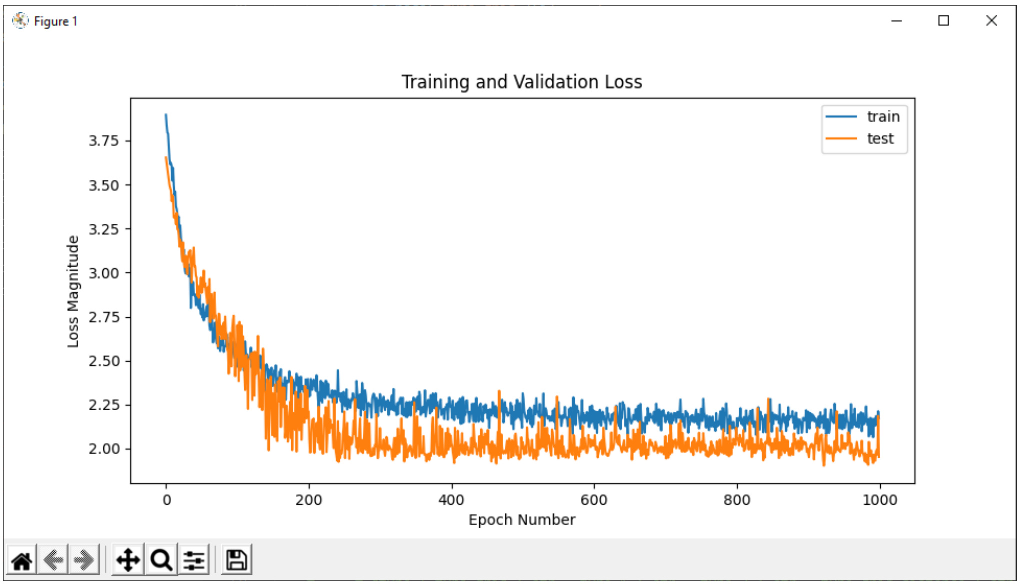
Firstly, we trained the model with 2016-2019 season data, and try to predict it on 2020 season races. The training and validation loss is very close to each other, indicating no overfitting or underfitting issue occurring. The result would be as follows:



Loss of training and testing, 2016-2019 seasons.

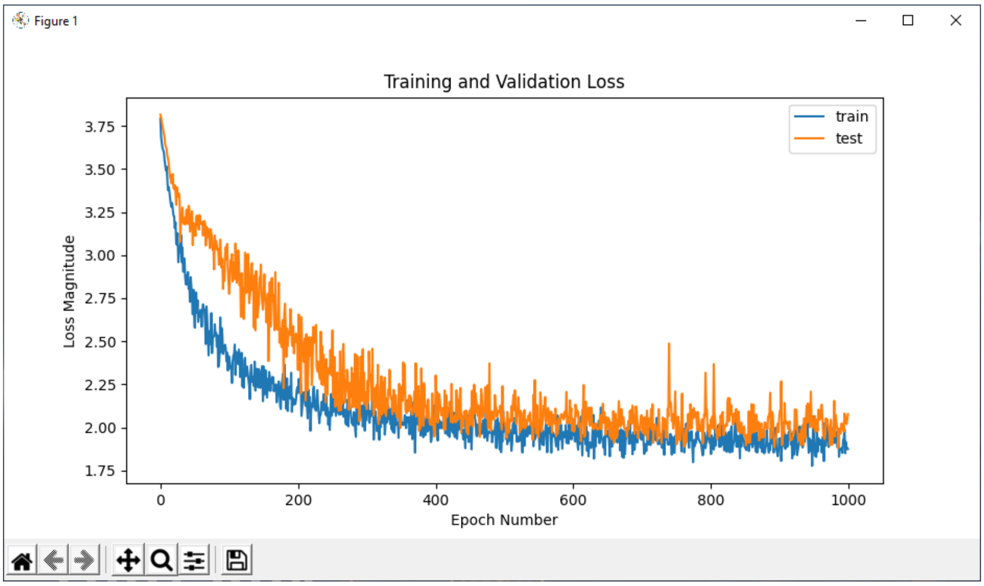
Name: Chan Ming Chung
UID: 3035373169

We also trained the model with 2016-2020 season data, but the loss of validation data is slightly lower than that of training data, which is not an ideal prediction model. The result would be as follows:



Loss of training and testing, 2016-2020 seasons.

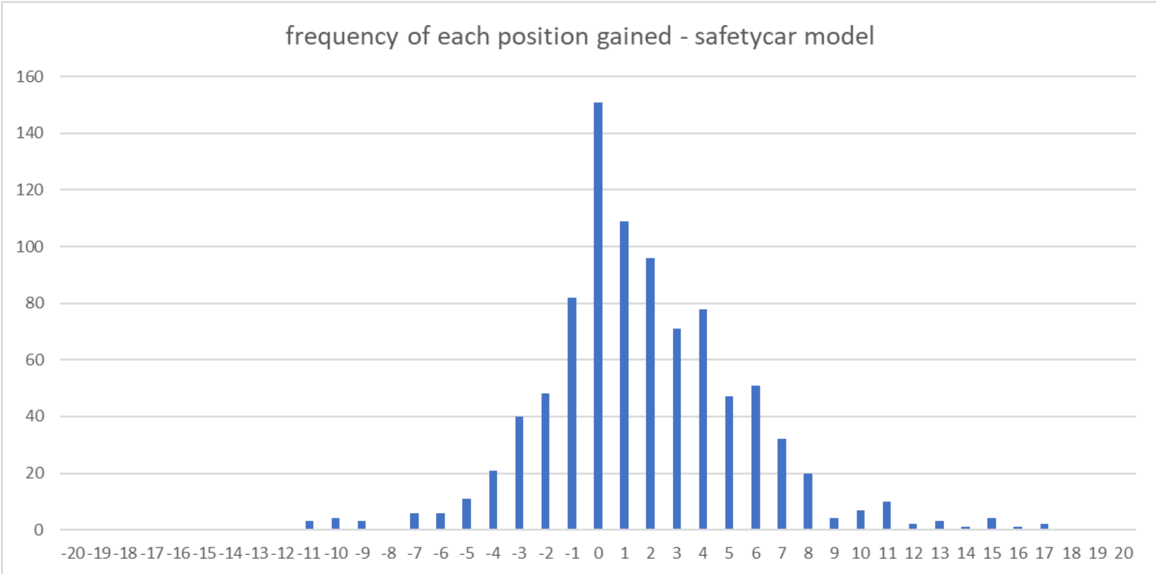
To predict the 2019 season, we have to train a model with 2016-2018 season data. The training data is slightly lower than that of validation data, which is also not ideal. The result would be as follows:



Loss of training and testing, 2016-2018 seasons.

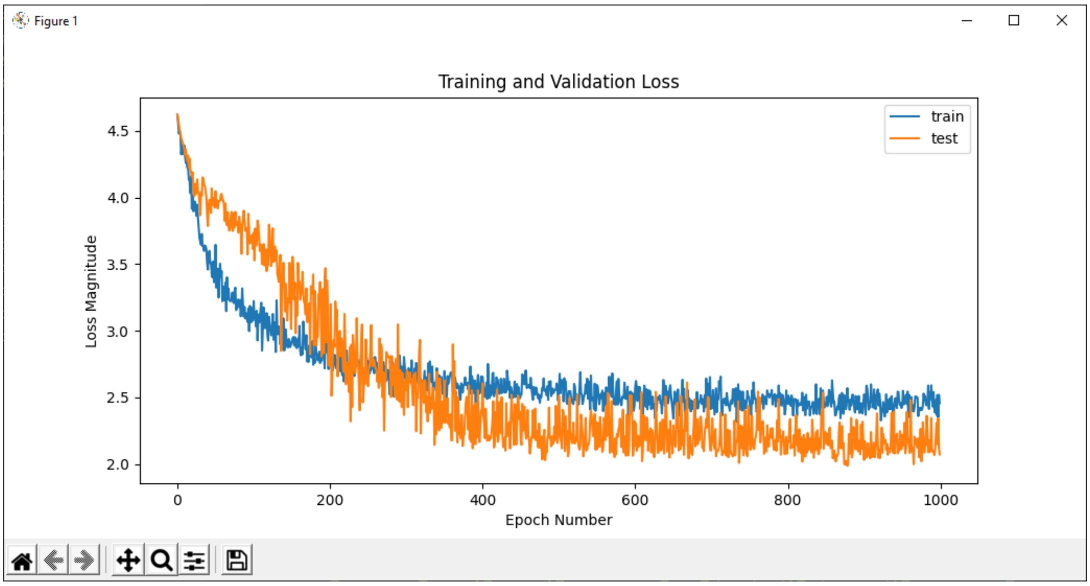
Name: Chan Ming Chung
UID: 3035373169

With the knowledge of model 1, we identified again that an imbalance of dataset result label exists that may have caused poor prediction results. The frequency of each position gained is shown in the following graph:



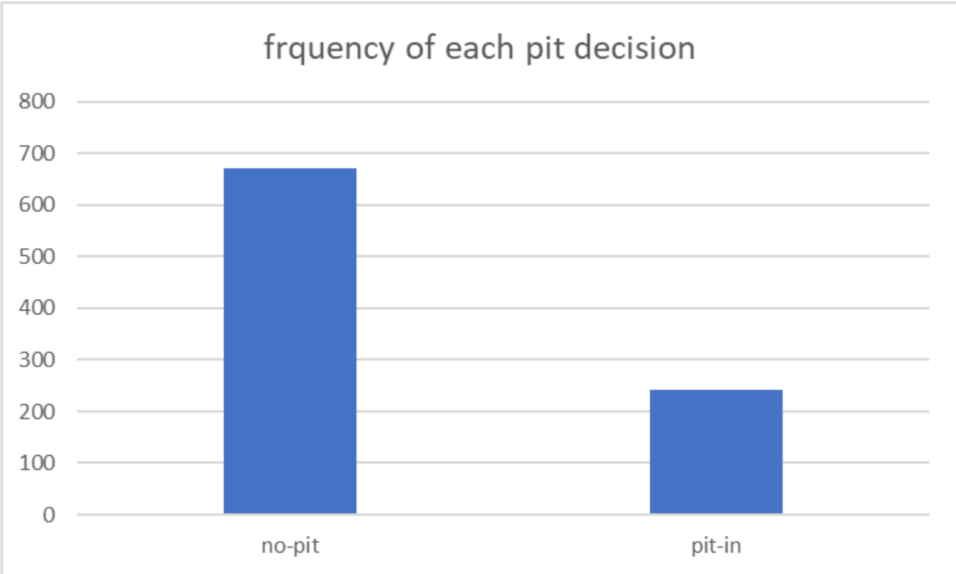
Frequency distribution of each position gained, 2016-2020 seasons.

We attempt to balance the dataset by cropping out high-frequency position gains, but the result showed that the validation data was worse than that of uncropped and tend to share similarity with the 2016-2020 seasons one. The result would be as follows:

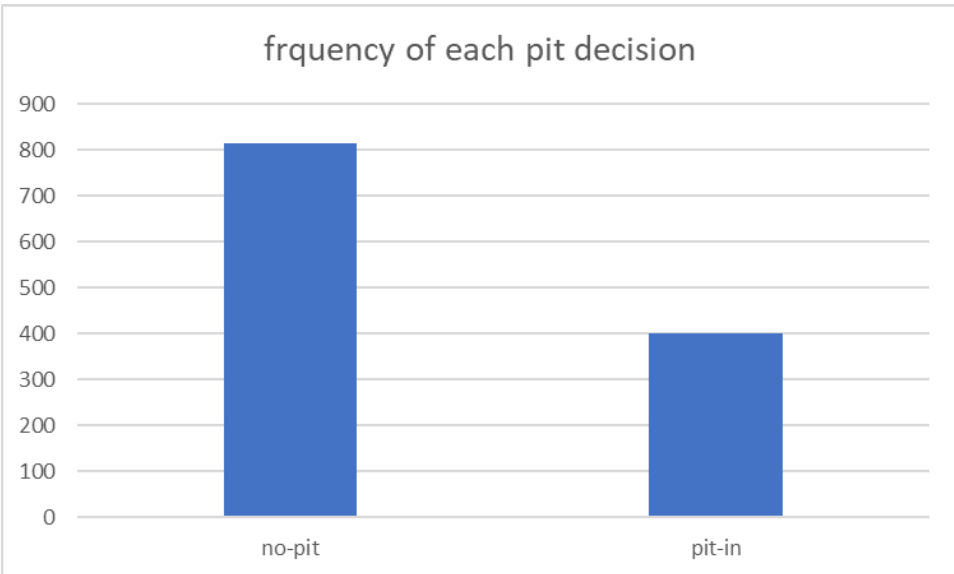


Loss of training and testing, high frequencies cropped, 2016-2019 seasons.

Since cropping out high-frequency data labels did not improve the model, we looked into the pit decision input column. Since this model focuses on the binary pitstop decision under safety car conditions, we investigated the distribution of the frequency of pit in and stay out. The result would be as follows:



Frequency of each pit decision (671:241), 2016-2019 seasons.



Frequency of each pit decision (813:401), 2016-2020 seasons.

From the results, we can see that 2016-2020 seasons had a “more even” distribution, about 2:1 ratio, while that of 2016-2019 was nearly 3:1. However, we already knew that the training and validation loss of 2016-2020 seasons was worse, therefore cropping the pit decision column data would also not improve the prediction accuracy.

Final Model design architecture

We located 58 races from 2016-2020 with mixed safety car decisions, which means some racers decided to pit in while some stays out during a safety car condition. For 2016-2019, 47 races falls into this criterion. The races are as follows:

2020 season: abudhabi2020, austrian2020, bahrain2020, belgian2020, british2020, eifel2020, imola2020, italian2020, russian2020, sakhir2020, tuscan2020

2019 season: azerbaijan2019, belgian2019, brazilian2019, british2019, chinese2019, italian2019, monaco2019, russian2019, singapore2019, spanish2019

2018 season: abudhabi2018, australian2018, austrian2018, azerbaijan2018, bahrain2018, belgian2018, british2018, canada2018, chinese2018, french2018, german2018, hungarian2018, italian2018, japanese2018, mexican2018, monaco2018, singapore2018, spanish2018, unitedstates2018

2017 season: azerbaijan2017, bahrain2017, belgian2017, brazilian2017, british2017, canada2017, hungarian2017, japanese2017, mexican2017, monaco2017, russian2017, spanish2017

2016 season: australian2016, austrian2016, belgian2016, canada2016, mexican2016, russian2016

For dividing the training data set and validation data set, we decided to split the data set by 20% of validation data randomly from the entire data source for each epoch during the training process. We adopted a similar model architecture as that of model 1, except that we have 15 input columns and 1 output column.

There were **1225 records from 2016-2020** season, and **923 records from 2016-2019 season**.

The model architecture would be as follows:

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
normalization (Normalization (None, 15)                31
_________________________________________________________________
dense (Dense)                (None, 256)               4096
_________________________________________________________________
dropout (Dropout)            (None, 256)               0
_________________________________________________________________
dense_1 (Dense)              (None, 256)               65792
_________________________________________________________________
dropout_1 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_2 (Dense)              (None, 128)               32896
_________________________________________________________________
dropout_2 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_3 (Dense)              (None, 128)               16512
_________________________________________________________________
dropout_3 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_4 (Dense)              (None, 64)                8256
_________________________________________________________________
dropout_4 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_5 (Dense)              (None, 64)                4160
_________________________________________________________________
dropout_5 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_6 (Dense)              (None, 32)                2080
_________________________________________________________________
dropout_6 (Dropout)          (None, 32)                0
_________________________________________________________________
dense_7 (Dense)              (None, 16)                528
_________________________________________________________________
dropout_7 (Dropout)          (None, 16)                0
_________________________________________________________________
dense_8 (Dense)              (None, 16)                272
_________________________________________________________________
dropout_8 (Dropout)          (None, 16)                0
_________________________________________________________________
dense_9 (Dense)              (None, 1)                 17
=================================================================
Total params: 134,640
Trainable params: 134,609
Non-trainable params: 31
```

Model 2: Final model design architecture.

Name: Chan Ming Chung
UID: 3035373169

Assessing the accuracy from the race result

The purpose of the second model is to predict the final position when deciding to go for a pitstop or stay out during a safety car scenario. We isolate the whole 2019 season and train / test the model with only the 2016-2018 seasons for the first test; Next, we isolate the whole 2020 season and train / test the model with only the 2016-2019 seasons. The results would be as follows:

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
39	1	30	12	3	8	0.002514043	-1	7	6.997485957	8
39	0	28	12	3	3	0.002508777	0	3	2.997491223	3
39	0	33	12	3	10	7.105106354	9	19	11.89489365	10
39	0	29	12	4	6	0.002505566	-1	5	4.997494434	6
39	0	5	12	4	5	2.460813999	3	8	5.539186001	5
39	0	29	12	3	9	2.957460642	4	13	10.04253936	9
39	1	32	12	3	13	0.065832227	-1	12	11.93416777	13
39	0	27	12	3	11	2.015784264	0	11	8.984215736	11
39	0	5	12	4	14	2.620761395	1	15	12.37923861	14
39	0	25	12	3	4	0.002514525	0	4	3.997485475	4
39	0	26	12	3	2	0.002504961	0	2	1.997495039	2
39	0	27	12	3	7	2.355892897	2	9	6.644107103	7
39	0	1	12	4	15	2.186313868	1	16	13.81368613	15
39	0	27	12	3	1	0.002503187	0	1	0.997496813	1
39	0	32	12	3	12	4.403099537	5	17	12.59690046	12

Azerbaijan 2019, accuracy index = 0.8667

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
1	1	1	43	3	17	0.00250334	-4	10	9.99749666	14
1	0	18	43	3	5	2.457234621	0	11	8.542765379	11
1	0	14	43	3	2	0.002509839	-2	2	1.997490161	4
1	1	1	43	3	18	0.002500661	-10	6	5.997499339	16
1	0	15	43	3	6	1.295123219	-4	9	7.704876781	13
1	0	12	43	3	9	2.882819653	4	13	10.11718035	9
1	0	14	43	3	8	0.002510346	1	7	6.997489654	6
1	0	20	43	3	1	0.002511627	0	1	0.997488373	1
1	0	13	43	3	10	3.722609758	6	16	12.27739024	10
1	0	24	43	3	7	0.943549931	-4	8	7.056450069	12
1	0	22	43	2	13	6.695579529	12	17	10.30442047	5
1	0	23	43	2	11	6.192342758	12	19	12.80765724	7
1	0	10	43	3	14	0.440998852	4	12	11.55900115	8
1	0	21	43	3	3	3.205302238	1	3	-0.205302238	2
1	0	29	43	2	15	2.811172724	-1	14	11.18882728	15
1	0	22	43	3	4	3.076747656	1	4	0.923252344	3
1	0	30	43	2	16	6.106963158	3	20	13.89303684	17
1	0	28	43	2	12	5.480028629	0	18	12.51997137	18

Belgian 2019, accuracy index = 3.5000

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
54	0	14	17	3	12	0.005401984	5	11	10.99459802	6
54	0	1	17	2	14	0.002506384	2	10	9.997493616	8
54	0	7	17	3	9	0.00251257	4	8	7.99748743	4
54	0	28	17	2	7	0.00250598	-6	7	6.99749402	13
54	0	7	17	3	6	0.425697118	4	6	5.574302882	2
54	0	9	17	3	11	2.125309467	6	15	12.87469053	9
54	0	1	17	3	16	0.002501453	-2	9	8.997498547	11
54	0	4	17	3	5	0.002507525	-9	5	4.997492475	14
54	0	8	17	3	13	2.162898779	6	16	13.83710122	10
54	0	25	17	2	8	5.38807106	17	20	14.61192894	3
54	0	8	17	3	10	2.100249767	7	12	9.899750233	5
66	0	26	5	3	8	1.878230572	5	11	9.121769428	6
66	0	13	5	2	9	1.940420389	2	10	8.059579611	8
66	0	19	5	3	6	0.849811673	4	8	7.150188327	4
66	0	1	5	3	15	0.002501139	-6	7	6.997498861	13
66	0	19	5	3	4	1.103790402	4	6	4.896209598	2
66	0	21	5	3	10	2.084117413	6	15	12.91588259	9
66	0	13	5	3	12	0.00250441	-2	9	8.99749559	11
66	0	16	5	3	2	0.002512628	-9	5	4.997487372	14
66	0	20	5	3	13	2.155636549	6	16	13.84436345	10
66	0	37	5	2	5	7.330506802	17	20	12.6694932	3
66	0	20	5	3	7	2.10223794	7	12	9.89776206	5

Brazilian 2019, accuracy index = 4.8182

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
20	1	8	32	2	8	0.002511085	0	7	6.997488915	7
20	0	7	32	2	9	0.002510352	-3	8	7.997489648	11
20	1	20	32	3	2	0.002509873	-10	6	5.997490127	16
20	0	3	32	1	14	1.120614767	4	12	10.87938523	8
20	0	8	32	1	7	0.00250423	1	5	4.99749577	4
20	0	2	32	1	12	3.750058889	-2	15	11.24994111	17
20	1	7	32	2	5	0.002501369	0	3	2.997498631	3
20	0	7	32	1	16	4.400392532	5	18	13.59960747	13
20	0	7	32	2	10	0.452733427	-3	9	8.547266573	12
20	1	8	32	2	13	3.381373882	8	17	13.61862612	9
20	0	7	32	1	11	1.108977079	0	10	8.891022921	10
20	1	7	32	2	4	0.002504143	-1	4	3.997495857	5
20	1	20	32	2	1	0.002502203	1	2	1.997497797	1
20	1	20	32	2	6	4.138921261	7	13	8.861078739	6
20	1	1	32	2	15	3.987228632	5	19	15.01277137	14
20	0	4	32	2	3	0.002500998	-1	1	0.997499002	2
20	1	20	32	2	17	4.40679884	5	20	15.59320116	15

British 2019, accuracy index = 2.4118

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
1	0	17	55	4	7	0.002507085	0	7	6.997492915	7
1	1	1	55	3	20	0.002506331	-3	15	14.99749367	18
1	0	17	55	3	4	0.002503034	0	3	2.997496966	3
1	0	24	55	3	12	0.002515431	4	13	12.99748457	9
1	0	7	55	4	10	0.002504685	-1	10	9.997495315	11
1	0	18	55	4	6	0.002508995	0	6	5.997491005	6
1	0	19	55	3	8	1.572831631	4	12	10.42716837	8
1	0	21	55	3	3	0.002505218	-1	4	3.997494782	5
1	0	19	55	3	13	0.501116991	4	16	15.49888301	12
1	0	8	55	4	11	0.002502987	-4	9	8.997497013	13
1	0	18	55	4	18	5.871065617	10	20	14.12893438	10
1	0	16	55	3	5	0.002508879	1	5	4.997491121	4
1	0	21	55	3	1	0.002502593	1	2	1.997497407	1
1	1	1	55	3	19	0.002510227	0	14	13.99748977	14
1	0	21	55	3	15	0.002512699	1	17	16.9974873	16
1	0	20	55	3	2	0.002501228	-1	1	0.997498772	2
1	0	25	55	3	16	0.154714435	1	18	17.84528556	17
1	0	6	55	4	14	4.837380409	4	19	14.16261959	15

Chinese 2019, accuracy index = 1.5556

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
29	1	1	24	4	3	0.002505259	1	5	4.997494741	4
29	0	6	24	2	11	4.142827511	6	16	11.85717249	10
29	0	3	24	4	19	5.91229105	5	20	14.08770895	15
29	1	22	24	3	18	0.002502643	-3	13	12.99749736	16
29	0	1	24	4	12	5.309952259	6	17	11.69004774	11
29	0	1	24	4	10	5.96007061	11	18	12.03992939	7
29	0	9	24	2	1	0.002501513	0	1	0.997498487	1
29	0	7	24	3	14	0.002501835	-3	9	8.997498165	12
29	0	3	24	3	8	0.00251376	2	8	7.99748624	6
29	1	29	24	4	5	0.002502053	1	6	5.997497947	5
29	0	2	24	3	13	8.610401154	11	19	10.38959885	8
29	0	10	24	3	2	0.002503559	-1	2	1.997496441	3
29	0	6	24	2	15	0.002511577	0	14	13.99748842	14
29	0	2	24	3	4	0.002503465	1	3	2.997496535	2
29	1	29	24	3	17	0.002505579	-2	15	14.99749442	17
29	0	8	24	3	7	1.93839705	1	10	8.06160295	9

Italian 2019, accuracy index = 1.6250

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
11	1	11	67	5	12	0.002501339	-3	6	5.997498661	9
11	0	36	67	4	10	1.640770674	1	12	10.35922933	11
11	1	11	67	5	4	0.002501489	2	4	3.997498511	2
11	0	35	67	5	13	0.002512815	-3	14	13.99748718	17
11	0	39	67	5	9	0.002511613	3	13	12.99748839	10
11	0	16	67	5	5	0.002506442	3	8	7.997493558	5
11	1	11	67	4	17	0.0025191	4	16	15.9974809	12
11	0	28	67	4	11	2.127199888	1	17	14.87280011	16
11	1	11	67	5	14	0.002500811	-9	5	4.997499189	14
11	0	29	67	5	8	0.002517553	2	10	9.997482447	8
11	0	21	67	5	7	0.002504225	0	7	6.997495775	7
11	0	2	67	3	18	0.002501254	-2	11	10.99749875	13
11	1	11	67	5	2	0.002501611	-1	3	2.997498389	4
11	1	11	67	5	1	0.002501091	0	1	0.997498909	1
11	0	19	67	5	6	0.002678117	3	9	8.997321883	6
11	0	1	67	4	19	0.002506932	4	19	18.99749307	15
11	2	11	67	5	3	0.002500891	-1	2	1.997499109	3
11	0	10	67	4	15	2.068849325	2	20	17.93115067	18
11	0	33	67	5	16	0.546302617	-1	18	17.45369738	19

Monaco 2019, accuracy index = 2.3158

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
1	0	19	52	4	6	0.002510934	-1	7	6.997489066	8
1	0	8	52	3	15	0.413266599	2	15	14.5867334	13
1	0	25	52	3	13	3.750070333	2	16	12.24992967	14
1	0	22	52	4	7	2.24539566	4	11	8.75460434	7
1	0	21	52	4	2	0.002501299	-2	1	0.997498701	3
1	0	26	52	3	11	2.74822402	3	14	11.25177598	11
1	0	26	52	4	10	2.58938098	4	13	10.41061902	9
1	0	28	52	3	18	7.467750549	15	20	12.53224945	5
1	0	26	52	2	14	4.67005682	7	19	14.32994318	12
1	0	15	52	4	9	0.00250355	-4	6	5.99749645	10
1	0	27	52	4	8	3.065203905	5	9	5.934796095	4
1	0	27	52	3	3	0.002504793	1	2	1.997495207	1
1	0	20	52	4	4	0.002508688	-1	5	4.997491312	6
1	0	27	52	3	5	0.00250734	2	4	3.99749266	2
1	1	1	52	4	12	0.002513109	-3	12	11.99748689	15
28	0	8	25	3	8	0.002508119	-1	7	6.997491881	8
28	0	1	25	4	15	0.877674043	2	15	14.12232596	13
28	0	2	25	4	13	2.805319071	2	16	13.19468093	14
28	0	5	25	3	9	1.54253459	4	11	9.45746541	7
28	1	6	25	3	3	0.002500862	-2	1	0.997499138	3
28	0	1	25	4	11	1.972541928	3	14	12.02745807	11
28	0	1	25	3	7	2.048267841	4	13	10.95173216	9
28	1	1	25	3	5	10.38045597	15	20	9.619544029	5
28	1	1	25	4	12	5.740386486	7	19	13.25961351	12
28	1	12	25	3	10	0.002503609	-4	6	5.997496391	10
28	1	28	25	4	4	4.228369713	5	9	4.771630287	4
28	1	28	25	3	1	0.002501355	1	2	1.997498645	1
28	0	7	25	3	6	0.002506384	-1	5	4.997493616	6
28	1	28	25	3	2	0.002502138	2	4	3.997497862	2
28	0	1	25	3	14	0.002506753	-3	12	11.99749325	15

Russian 2019, accuracy index = 2.0333

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
36	0	2	25	3	16	3.914054632	6	20	16.08594537	14
36	0	16	25	3	7	0.00250623	2	9	8.99749377	7
36	0	17	25	3	1	0.002501419	2	3	2.997498581	1
36	0	1	25	5	18	0.002511347	6	17	16.99748865	11
36	0	4	25	4	13	0.002505203	3	11	10.9974948	8
36	0	16	25	3	2	0.002501093	-1	1	0.997498907	2
36	0	18	25	3	8	1.533977032	-4	13	11.46602297	17
36	0	16	25	3	6	0.00250233	0	6	5.99749767	6
36	1	24	25	3	14	1.877454281	-1	14	12.12254572	15
36	1	35	25	3	9	0.002515168	-1	8	7.997484832	9
36	0	17	25	3	3	0.002502584	1	4	3.997497416	3
36	0	10	25	3	4	0.002501122	-2	2	1.997498878	4
36	0	1	25	4	19	0.00250108	-5	7	6.99749892	12
36	0	14	25	3	5	0.002501476	0	5	4.997498524	5
36	0	15	25	3	17	1.265540361	3	19	17.73445964	16
36	1	2	25	3	15	0.002502557	0	10	9.997497443	10
44	0	10	17	3	14	4.993338585	6	20	15.00666142	14
44	0	24	17	3	7	0.002509852	2	9	8.997490148	7
44	0	25	17	3	1	0.002501598	2	3	2.997498402	1
44	0	9	17	5	15	1.946834207	6	17	15.05316579	11
44	0	12	17	4	10	0.002511632	3	11	10.99748837	8
44	0	24	17	3	2	0.00250117	-1	1	0.99749883	2
44	0	26	17	3	8	2.94291687	-4	13	10.05708313	17
44	0	24	17	3	6	0.002503245	0	6	5.997496755	6
44	0	8	17	5	12	0.336123586	-1	14	13.66387641	15
44	0	8	17	4	11	0.002502435	-1	8	7.997497565	9
44	0	25	17	3	3	0.002503352	1	4	3.997496648	3
44	0	18	17	3	4	0.002501208	-2	2	1.997498792	4
44	0	9	17	4	17	0.002501336	-5	7	6.997498664	12
44	0	22	17	3	5	0.002502266	0	5	4.997497734	5
44	1	23	17	3	16	5.337033749	3	19	13.66296625	16
44	0	8	17	5	13	0.002503178	0	10	9.997496822	10

Singapore 2019, accuracy index = 2.1250

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
46	0	1	20	2	13	2.334659338	1	13	10.66534066	12
46	0	6	20	2	5	0.002508494	-1	3	2.997491506	4
46	0	1	20	3	15	1.695783138	0	14	12.30421686	14
46	1	20	20	2	7	0.002509874	-3	7	6.997490126	10
46	1	24	20	2	6	0.002509024	0	6	5.997490976	6
46	0	1	20	3	14	2.38883996	0	15	12.61116004	15
46	1	21	20	1	3	0.002509616	0	5	4.997490384	5
46	0	1	20	3	8	0.0026968	1	8	7.9973032	7
46	0	1	20	2	11	1.978206396	0	11	9.021793604	11
46	0	1	20	2	9	1.504109263	0	9	7.495890737	9
46	0	11	20	3	12	4.156020641	7	20	15.84397936	13
46	0	3	20	2	4	1.38904655	1	4	2.61095345	3
46	1	19	20	2	1	0.002502204	1	2	1.997497796	1
46	1	22	20	2	10	2.180505276	4	12	9.819494724	8
46	0	2	20	2	17	3.664895296	2	19	15.3351047	17
46	0	1	20	3	2	0.002502798	-1	1	0.997497202	2
46	0	2	20	2	18	1.433791637	-1	17	15.56620836	18
46	0	5	20	3	16	2.712022781	2	18	15.28797722	16

Spanish 2019, accuracy index = 1.4444

In the 2019 season, the total average accuracy index is 2.2696, which in other words the final position prediction would have ±2.2696 error.

For the 2020 season, we included 2019 season data as well in the training process, and the results are as follows:

Name: Chan Ming Chung
UID: 3035373169

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
10	0	29	46	3	10	1.870329022	4	11	9.129670978	7
10	1	10	46	5	5	-0.001855818	-1	4	4.001855818	5
10	0	25	46	3	11	1.785262346	-1	13	11.21473765	14
10	1	1	46	4	19	0.108784251	1	18	17.89121575	17
10	1	10	46	4	14	1.595532179	3	15	13.40446782	12
10	1	10	46	5	8	1.624040604	1	9	7.375959396	8
10	0	12	46	4	13	1.719146729	-1	12	10.28085327	13
10	1	10	46	5	7	1.017808318	-2	8	6.982191682	10
10	0	21	46	3	17	2.504166842	2	20	17.49583316	18
10	1	10	46	5	4	-0.001855818	1	5	5.001855818	4
10	1	10	46	5	9	-0.001855818	-4	7	7.001855818	11
10	1	10	46	4	12	-0.001855818	1	10	10.00185582	9
10	1	10	46	4	1	-0.001855818	0	1	1.001855818	1
10	1	10	46	4	3	-0.001855818	0	3	3.001855818	3
10	1	10	46	4	6	-0.001855818	0	6	6.001855818	6
10	1	10	46	4	16	1.382460237	1	16	14.61753976	15
10	1	10	46	4	2	-0.001855818	0	2	2.001855818	2
10	0	17	46	4	15	-0.001473691	-2	14	14.00147369	16

Abu Dhabi 2020, accuracy index = 1.3333

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
26	1	26	45	4	1	-0.001855818	0	1	1.001855818	1
51	1	25	20	2	1	-0.001855818	0	1	1.001855818	1
26	1	26	45	4	4	-0.001855818	0	3	3.001855818	3
51	1	25	20	2	5	-0.001855818	0	3	3.001855818	3
26	1	26	45	4	2	1.819190264	1	5	3.180809736	4
51	1	25	20	2	2	1.462686777	1	5	3.537313223	4
26	1	26	45	4	3	-0.001855818	-9	4	4.001855818	13
51	1	25	20	2	3	-0.001855818	-9	4	4.001855818	13
51	1	25	20	3	4	1.016976833	0	6	4.983023167	6
26	1	26	45	4	5	-0.001855818	0	6	6.001855818	6
51	1	25	20	2	6	0.150273755	5	7	6.849726245	2
26	1	26	45	4	6	-0.001855818	5	7	7.001855818	2
51	1	25	20	2	7	0.01485251	3	8	7.98514749	5
26	1	26	45	4	7	-0.001855818	3	8	8.001855818	5
51	1	25	20	2	8	2.830762625	5	12	9.169237375	7
26	1	26	45	3	8	1.716627955	1	11	9.283372045	10
26	1	26	45	3	9	2.010626793	5	12	9.989373207	7
51	1	25	20	2	9	2.662598372	1	13	10.33740163	12
51	1	25	20	2	13	-0.001855818	1	11	11.00185582	10
26	1	26	45	3	10	1.741404653	1	13	11.25859535	12
51	1	25	20	2	10	2.184203148	6	14	11.81579685	8
26	1	26	45	3	11	0.89873457	6	14	13.10126543	8
26	0	1	45	2	12	4.624022484	9	18	13.37597752	9
51	1	26	20	2	11	3.655163527	9	18	14.34483647	9
51	1	25	20	2	14	3.586251974	9	20	16.41374803	11
26	1	26	45	3	14	3.294052362	9	20	16.70594764	11

Austrian 2020, accuracy index = 2.8846

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
1	1	15	57	3	5	-0.001855818	-1	6	6.001855818	7
1	1	17	57	3	9	-0.001855818	5	9	9.001855818	4
1	1	18	57	2	15	-0.001855818	-2	11	11.00185582	13
1	1	19	57	3	16	3.107023478	6	20	16.89297652	14
1	1	15	57	3	17	0.587009311	2	17	16.41299069	15
1	1	23	57	2	7	-0.001855818	2	8	8.001855818	6
1	1	18	57	3	3	-0.001855818	-13	5	5.001855818	18
1	1	21	57	3	10	1.413740277	2	12	10.58625972	10
1	1	17	57	3	4	-0.001855818	1	4	4.001855818	3
1	1	15	57	4	19	-0.001855818	-1	10	10.00185582	11
1	1	15	57	3	8	-0.001855818	-2	7	7.001855818	9
1	1	18	57	3	1	-0.001855818	0	1	1.001855818	1
1	1	19	57	4	13	2.644310951	10	15	12.35568905	5
1	1	18	57	3	18	-0.001855818	2	14	14.00185582	12
1	1	20	57	2	6	-0.001855818	-6	2	2.001855818	8
1	1	18	57	3	11	1.430640697	0	16	14.5693593	16
55	0	19	3	2	7	-0.001855818	-1	6	6.001855818	7
55	0	17	3	2	4	1.816727281	5	9	7.183272719	4
55	0	31	3	2	6	1.760388255	2	8	6.239611745	6
55	0	22	3	2	3	-0.001855818	1	4	4.001855818	3
55	0	22	3	2	9	-0.001855818	-2	7	7.001855818	9
55	0	20	3	2	1	-0.001855818	0	1	1.001855818	1
55	0	17	3	2	5	5.252140045	10	15	9.747859955	5

Bahrain 2020, accuracy index = 2.6087

Name: Chan Ming Chung
UID: 3035373169

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
10	1	1	35	4	4	-0.001855818	0	4	4.001855818	4
10	1	1	35	4	10	1.38516748	3	10	8.61483252	7
10	1	1	35	3	12	1.128527045	1	14	12.87147295	13
10	1	10	35	3	17	2.025309086	3	19	16.97469091	16
10	1	10	35	3	14	1.620860696	4	16	14.3791393	12
10	1	10	35	3	15	1.711027265	2	17	15.28897274	15
10	0	16	35	2	8	2.536876678	4	12	9.463123322	8
10	0	8	35	4	9	2.086961508	-2	8	5.913038492	10
10	1	10	35	4	13	1.889680862	-1	13	11.11031914	14
10	1	1	35	4	7	1.388378143	0	9	7.611621857	9
10	1	10	35	3	16	2.610236883	3	20	17.38976312	17
10	1	1	35	4	6	-0.001855818	-1	5	5.001855818	6
10	1	1	35	3	11	0.755848169	0	11	10.24415183	11
10	1	1	35	4	5	-0.001855818	1	6	6.001855818	5
10	1	1	35	3	3	-0.001855818	0	3	3.001855818	3
10	1	1	35	3	1	-0.001855818	0	1	1.001855818	1
10	1	1	35	3	2	-0.001855818	0	2	2.001855818	2

Belgian 2020, accuracy index = 1.0000

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
2	0	10	51	3	6	-0.001855818	4	8	8.001855818	4
2	0	10	51	3	7	-0.001855818	0	5	5.001855818	5
2	0	10	51	3	10	-0.001855818	0	10	10.00185582	10
2	0	10	51	2	18	2.271079779	3	18	15.72892022	15
2	0	10	51	2	16	2.194471598	-1	16	13.8055284	17
2	0	34	51	2	15	2.612563848	1	17	14.38743615	16
2	0	10	51	2	11	-0.001855818	4	11	11.00185582	7
2	0	11	51	2	4	-0.001855818	1	4	4.001855818	3
2	0	11	51	2	8	-0.001855818	-3	6	6.001855818	9
2	0	4	51	2	12	-0.001855818	4	12	12.00185582	8
2	0	10	51	3	9	-0.001855818	3	9	9.001855818	6
2	0	11	51	2	3	-0.001855818	1	3	3.001855818	2
2	0	11	51	2	1	-0.001855818	0	1	1.001855818	1
2	0	11	51	3	5	-0.001855818	-6	7	7.001855818	13
2	0	10	51	2	17	2.707284212	8	20	17.29271579	12
2	0	11	51	2	2	-0.001855818	-9	2	2.001855818	11
2	0	10	51	2	13	1.041894794	1	15	13.95810521	14
12	1	12	41	3	5	1.519884825	4	8	6.480115175	4
12	1	12	41	3	7	-0.001855818	0	5	5.001855818	5
12	1	12	41	3	10	0.598660946	0	10	9.401339054	10
12	1	12	41	2	16	2.440600157	3	18	15.55939984	15
12	1	12	41	2	13	2.841792583	-1	16	13.15820742	17
12	0	24	41	2	14	2.375966787	1	17	14.62403321	16
12	1	12	41	2	11	1.853173256	4	11	9.146826744	7
12	1	1	41	2	4	-0.001855818	1	4	4.001855818	3
12	1	1	41	2	8	-0.001855818	-3	6	6.001855818	9
12	0	6	41	1	17	-0.001855818	4	12	12.00185582	8
12	1	12	41	3	9	-0.001855818	3	9	9.001855818	6
12	1	1	41	2	3	-0.001855818	1	3	3.001855818	2
12	1	1	41	2	1	-0.001855818	0	1	1.001855818	1
12	1	1	41	3	6	-0.001855818	-6	7	7.001855818	13
12	1	12	41	2	15	4.610683441	8	20	15.38931656	12
12	1	1	41	2	2	-0.001855818	-9	2	2.001855818	11
12	1	12	41	2	12	2.663425922	1	15	12.33657408	14

British 2020, accuracy index = 2.6176

Name: Chan Ming Chung
UID: 3035373169

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
16	1	16	45	4	4	2.159947634	3	6	3.840052366	3
16	0	5	45	2	16	-0.001855818	0	11	11.00185582	11
16	0	2	45	3	19	-0.001855818	4	18	18.00185582	14
16	0	6	45	3	17	1.996094108	7	19	17.00390589	12
16	0	12	45	3	14	1.800216317	7	16	14.19978368	9
16	0	14	45	3	9	2.532086134	6	12	9.467913866	6
16	0	12	45	4	6	2.356525183	5	9	6.643474817	4
16	0	6	45	3	12	-0.001855818	-3	4	4.001855818	7
16	0	2	45	3	18	-0.001855818	2	15	15.00185582	13
16	1	1	45	3	13	-0.001855818	-2	13	13.00185582	15
16	0	13	45	4	11	5.279485226	12	20	14.72051477	8
16	1	16	45	4	2	0.882591367	1	3	2.117408633	2
16	1	16	45	4	1	0.048182059	1	2	1.951817941	1
16	0	12	45	4	8	2.106930256	5	10	7.893069744	5
16	0	1	45	3	15	-0.001855818	4	14	14.00185582	10
45	0	1	16	4	4	2.782906532	3	6	3.217093468	3
45	0	4	16	4	13	-0.001855818	0	11	11.00185582	11
45	0	12	16	3	12	3.469505548	4	18	14.53049445	14
45	0	1	16	4	14	3.11202836	7	19	15.88797164	12
45	0	17	16	2	7	5.314285755	7	16	10.68571424	9
45	0	1	16	4	8	3.146031618	6	12	8.853968382	6
45	1	17	16	3	3	4.326954842	5	9	4.673045158	4
45	0	10	16	3	6	-0.001855818	-3	4	4.001855818	7
45	0	11	16	3	11	2.805055618	2	15	12.19494438	13
45	0	1	16	4	15	-0.001855818	-2	13	13.00185582	15
45	0	1	16	4	9	5.745611668	12	20	14.25438833	8
45	1	29	16	3	2	1.490655303	1	3	1.509344697	2
45	1	29	16	3	1	0.227288678	1	2	1.772711322	1
45	0	1	16	4	5	3.34526515	5	10	6.65473485	5
45	0	9	16	3	10	2.876338959	4	14	11.12366104	10

Eifel 2020, accuracy index = 2.3333

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
30	0	16	34	2	8	-0.001855818	2	5	5.001855818	3
30	0	17	34	3	13	-0.001855818	1	9	9.001855818	8
30	0	9	34	3	4	3.315137148	2	14	10.68486285	12
30	0	4	34	3	6	5.518064022	8	19	13.48193598	11
30	0	18	34	3	5	5.329478264	9	18	12.67052174	9
30	0	21	34	2	16	1.833191037	2	16	14.16680896	14
30	0	3	34	2	7	1.410389304	5	11	9.589610696	6
30	0	17	34	2	9	0.736422718	2	7	6.263577282	5
30	0	16	34	2	10	-0.001855818	-9	6	6.001855818	15
30	0	16	34	2	11	-0.001855818	4	8	8.001855818	4
30	1	30	34	3	1	-0.001855818	1	2	2.001855818	1
30	0	13	34	3	12	1.30418098	3	10	8.69581902	7
30	0	11	34	2	2	-0.001855818	-1	1	1.001855818	2
30	0	20	34	3	15	2.776606798	10	20	17.2233932	10
51	0	37	13	2	4	1.285396934	2	5	3.714603066	3
51	1	38	13	3	7	2.158453465	1	9	6.841546535	8
51	1	12	13	2	14	2.07770896	2	14	11.92229104	12
51	0	17	13	2	13	3.568277597	8	19	15.4317224	11
51	0	3	13	4	11	3.492260695	9	18	14.50773931	9
51	0	1	13	4	16	0.762788594	2	16	15.23721141	14
51	1	24	13	2	3	4.438999176	5	11	6.561000824	6
51	0	38	13	2	5	1.415999532	2	7	5.584000468	5
51	0	37	13	2	6	0.457632691	-9	6	5.542367309	15
51	1	37	13	2	8	-0.001855818	4	8	8.001855818	4
51	1	21	13	2	1	-0.001855818	1	2	2.001855818	1
51	1	34	13	3	9	1.461517334	3	10	8.538482666	7
51	1	32	13	2	2	-0.001855818	-1	1	1.001855818	2
51	0	41	13	3	12	5.469810009	10	20	14.53018999	10

Imola 2020, accuracy index = 2.7143

Name: Chan Ming Chung
UID: 3035373169

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
20	1	2	34	4	5	0.520615697	1	7	6.479384303	6
20	1	2	34	4	3	-0.001855818	2	6	6.001855818	4
20	0	4	34	2	18	3.02980566	9	20	16.97019434	11
20	0	2	34	2	16	-0.001855818	1	14	14.00185582	13
20	1	2	34	3	12	2.225082159	4	16	13.77491784	12
20	0	1	34	2	15	-0.001855818	9	10	10.00185582	1
20	1	2	34	4	4	-0.001855818	-6	4	4.001855818	10
20	0	6	34	4	8	-0.001855818	5	8	8.001855818	3
20	1	2	34	4	11	-0.001855818	-6	9	9.001855818	15
20	1	2	34	2	10	0.070147052	2	11	10.92985295	9
20	1	2	34	4	9	1.535165429	4	12	10.46483457	8
20	1	20	34	4	1	-0.001855818	-6	1	1.001855818	7
20	1	2	34	4	2	-0.001855818	1	3	3.001855818	2
20	1	2	34	3	14	3.216108084	5	19	15.78389192	14
20	1	2	34	4	6	-0.001855818	-3	2	2.001855818	5
20	1	20	34	3	13	2.617656469	2	18	15.38234353	16
25	1	3	29	3	10	-0.001855818	1	7	7.001855818	6
25	1	3	29	3	7	-0.001855818	2	6	6.001855818	4
25	1	9	29	2	9	6.750068665	9	20	13.24993134	11
25	1	7	29	2	4	5.724299431	1	14	8.275700569	13
25	1	3	29	2	17	0.000163352	4	16	15.99983665	12
25	1	6	29	2	3	4.27096796	9	10	5.72903204	1
25	1	3	29	3	14	-0.001855818	-6	4	4.001855818	10
25	1	1	29	4	2	2.707167387	5	8	5.292832613	3
25	1	3	29	3	16	-0.001855818	-6	9	9.001855818	15
25	1	3	29	3	13	-0.001855818	2	11	11.00185582	9
25	1	3	29	3	12	-0.001855818	4	12	12.00185582	8
25	1	5	29	3	1	-0.001855818	-6	1	1.001855818	7
25	1	3	29	3	6	-0.001855818	1	3	3.001855818	2
25	1	3	29	2	15	3.169024229	5	19	15.83097577	14
25	1	3	29	3	8	-0.001855818	-3	2	2.001855818	5
25	1	5	29	2	5	7.799109936	2	18	10.20089006	16

Italian 2020, accuracy index = 3.4688

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
1	0	14	53	5	5	-0.001855818	0	5	5.001855818	5
1	1	1	53	5	18	-0.001855818	-7	8	8.001855818	15
1	0	29	53	4	13	3.0321908	1	14	10.9678092	13
1	0	16	53	4	14	4.872763634	4	20	15.12723637	16
1	0	34	53	3	15	4.302609921	5	19	14.69739008	14
1	0	16	53	4	10	2.876669407	-1	16	13.12333059	17
1	0	17	53	5	7	-0.001855818	0	9	9.001855818	9
1	0	19	53	5	6	-0.001855818	0	4	4.001855818	4
1	0	27	53	4	8	2.212415934	4	10	7.787584066	6
1	0	18	53	4	9	4.634405136	6	18	13.36559486	12
1	1	1	53	5	17	5.083414555	5	15	9.916585445	10
1	0	29	53	3	11	2.281761885	3	11	8.718238115	8
1	0	17	53	5	4	-0.001855818	0	7	7.001855818	7
1	0	24	53	4	3	-0.001855818	0	2	2.001855818	2
1	0	15	53	5	1	-0.001855818	-2	1	1.001855818	3
1	0	25	53	4	2	-0.001855818	2	3	3.001855818	1
1	0	15	53	4	12	3.016448736	6	17	13.98355126	11
42	0	27	12	3	5	-0.001855818	0	5	5.001855818	5
42	0	41	12	3	10	-0.001855818	-7	8	8.001855818	15
42	0	12	12	3	14	1.573681951	1	14	12.42631805	13
42	0	25	12	3	17	3.659117699	4	20	16.3408823	16
42	0	7	12	4	15	3.622654438	5	19	15.37734556	14
42	1	25	12	3	16	-0.001855818	-1	16	16.00185582	17
42	1	24	12	3	9	-0.001855818	0	9	9.001855818	9
42	0	22	12	3	4	-0.001855818	0	4	4.001855818	4
42	0	14	12	3	6	1.76425302	4	10	8.23574698	6
42	0	23	12	3	12	3.723710775	6	18	14.27628922	12
42	0	15	12	4	11	3.844974995	5	15	11.15502501	10
42	0	12	12	4	8	1.789934397	3	11	9.210065603	8
42	0	24	12	3	7	-0.001855818	0	7	7.001855818	7
42	0	17	12	3	2	-0.001855818	0	2	2.001855818	2
42	0	26	12	3	3	-0.001855818	-2	1	1.001855818	3
42	0	16	12	3	1	-0.001855818	2	3	3.001855818	1
42	0	26	12	3	13	3.051690578	6	17	13.94830942	11

Russian 2020, accuracy index = 1.4412

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
1	0	28	87	4	4	0.136820987	2	7	6.863179013	5
1	0	19	87	4	10	6.714448452	9	19	12.28555155	10
1	0	27	87	4	7	-0.001855818	-2	9	9.001855818	11
1	1	1	87	4	18	-0.001855818	4	5	5.001855818	1
1	0	41	87	4	6	-0.001855818	7	10	10.00185582	3
1	0	25	87	4	11	-0.001855818	0	15	15.00185582	15
1	0	46	87	3	12	1.22191143	6	12	10.77808857	6
1	0	26	87	4	5	-0.001855818	-1	6	6.001855818	7
1	0	40	87	3	8	-0.001855818	9	11	11.00185582	2
1	0	27	87	4	3	1.100643635	4	8	6.899356365	4
1	0	48	87	3	2	-0.001855818	-7	1	1.001855818	8
1	0	25	87	3	13	-0.001855818	1	14	14.00185582	13
55	1	26	33	3	4	3.191271544	2	7	3.808728456	5
55	0	1	33	3	11	6.787454605	9	19	12.21254539	10
55	0	4	33	2	10	3.317566633	-2	9	5.682433367	11
55	0	8	33	2	7	1.267551541	4	5	3.732448459	1
55	0	13	33	3	6	3.225856066	7	10	6.774143934	3
55	0	1	33	4	14	4.080880165	0	15	10.91911983	15
55	0	8	33	2	8	3.78055954	6	12	8.21944046	6
55	0	2	33	2	9	-0.001855818	-1	6	6.001855818	7
55	0	14	33	2	5	3.835765123	9	11	7.164234877	2
55	1	27	33	3	3	3.504583359	4	8	4.495416641	4
55	0	6	33	2	2	-0.001855818	-7	1	1.001855818	8
55	0	1	33	3	13	4.110684872	1	14	9.889315128	13
61	0	6	27	2	8	2.819227457	2	7	4.180772543	5
61	0	7	27	3	11	6.498506069	9	19	12.50149393	10
61	0	10	27	2	10	2.903251171	-2	9	6.096748829	11
61	0	14	27	2	3	2.359274626	4	5	2.640725374	1
61	0	19	27	3	5	3.282975674	7	10	6.717024326	3
61	0	7	27	4	14	3.633578062	0	15	11.36642194	15
61	1	14	27	2	6	4.873645306	6	12	7.126354694	6
61	0	8	27	2	9	-0.001855818	-1	6	6.001855818	7
61	0	20	27	2	4	3.95458293	9	11	7.04541707	2
61	0	6	27	3	7	3.129373789	4	8	4.870626211	4
61	1	12	27	2	2	-0.001855818	-7	1	1.001855818	8
61	0	7	27	3	13	3.600703478	1	14	10.39929652	13

Sakhir 2020, accuracy index = 3.2500

sc_lap	sc_decision	sc_laps_travelled	sc_laps_remaining	sc_tyre_compound	before_pit_pos	prediction	answer_final_pos_gained	initial_position	predicted_final_pos	answer_final_pos
1	1	7	59	3	6	0.071181037	4	8	7.928818963	4
1	1	7	59	3	8	2.182218313	5	11	8.817781687	6
1	1	1	59	3	18	-0.001855818	4	14	14.00185582	10
1	1	1	59	2	16	0.666984797	4	13	12.3330152	9
1	1	7	59	2	17	1.926896572	3	15	13.07310343	12
1	1	7	59	3	7	-0.001855818	2	7	7.001855818	5
1	1	7	59	3	3	-0.001855818	-3	5	5.001855818	8
1	1	7	59	3	4	-0.001855818	1	4	4.001855818	3
1	1	7	59	2	9	2.333876133	5	12	9.666123867	7
1	1	7	59	3	2	-0.001855818	0	1	1.001855818	1
1	1	7	59	2	11	4.649836063	7	18	13.35016394	11
1	1	7	59	3	1	-0.001855818	0	2	2.001855818	2
44	1	1	16	3	3	3.517616034	4	8	4.482383966	4
44	1	1	16	3	6	3.350751162	5	11	7.649248838	6
44	1	1	16	3	10	2.691608667	4	14	11.30839133	10
44	1	0	16	1	12	2.773162127	4	13	10.22683787	9
44	1	0	16	3	11	2.792485476	3	15	12.20751452	12
44	1	1	16	3	5	1.363778114	2	7	5.636221886	5
44	1	7	16	2	8	-0.001855818	-3	5	5.001855818	8
44	1	1	16	3	4	-0.001855818	1	4	4.001855818	3
44	1	1	16	3	7	3.027392149	5	12	8.972607851	7
44	1	0	16	1	1	-0.001855818	0	1	1.001855818	1
44	1	1	16	3	9	5.775616169	7	18	12.22438383	11
44	1	1	16	2	2	-0.001855818	0	2	2.001855818	2

Tuscan 2020, accuracy index = 1.5833

In the 2020 season, the total average accuracy index is 2.0999, which in other words the final position prediction would have ±2.0999 error.

We would like to know if the model improves its accuracy or not by including more years of data, so we attempt to train the model with several different sets of data from 2016-2019, by adding 1 race for each generation of model to see how the accuracy index changes when we increase the number of races to be trained. We test all models by using the races in the 2020 season. The accuracy index results are simplified as follows:

2016-2019 full seasons: 2.0999

2016-2018 + first 9 races in 2019: 2.2651

2016-2018 + first 8 races in 2019: 2.1538

2016-2018 + first 7 races in 2019: 2.1212

2016-2018 + first 6 races in 2019: 2.2518

2016-2018 + first 5 races in 2019: 2.1993

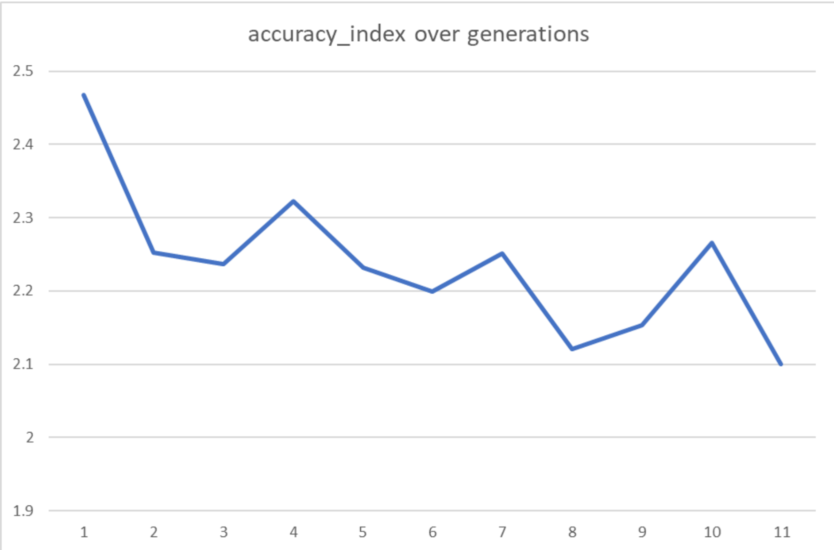
2016-2018 + first 4 races in 2019: 2.2321

2016-2018 + first 3 races in 2019: 2.3220

2016-2018 + first 2 races in 2019: 2.2362

2016-2018 + first race in 2019: 2.2521

2016-2018: 2.4674



Graph representation of accuracy index over generations

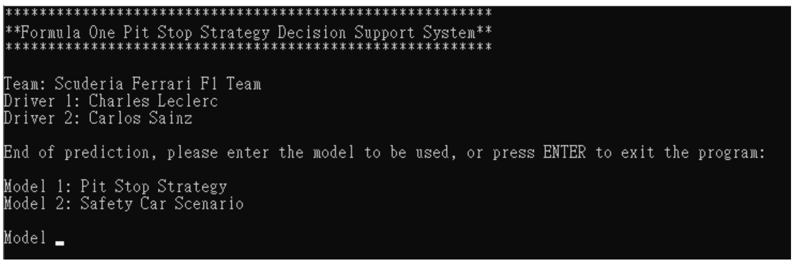
The result shows that the first model that was trained with 2016-2018 seasons have ±2.4674 final position difference in 2020 season, which is higher than ±2.0999 of the last model trained with 2016-2019 season. There is also an observable decreasing trend of the accuracy index from the first to the last model. We may conclude that including more seasons in the future would improve the accuracy of the model.

Decision Support System

To demonstrate how the race strategist in the Scuderia Ferrari F1 Team could use the model to swiftly come up with a pitstop strategy decision, we make a simplified DSS that provides a real-time prediction function by accessing the two established models, which are predicting final position gain or loss, and predicting pitstop strategy under Safety Car conditions, respectively.

We provide a simple Python program with a command line interface, with guidance that leads the user to input the predicting team, racer, and the corresponding required race information. The user is only required to use the numeric keypad to navigate through the DSS, such that the least amount of typing is required to arrive at the wanted predictions.

The user interface is shown in the following examples:



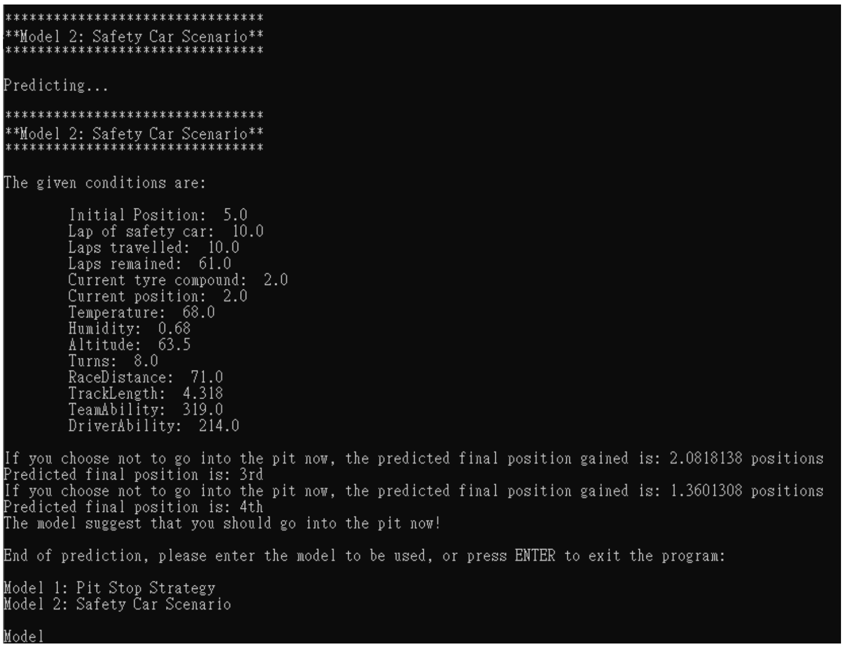
DSS UI, asking which model to be used.

```
******************************
**Model 1: Pit Stop Strategy**
******************************

Please select the circuit

        1. Abu Dhabi: Yas Marina
        2. Australian: Albert Park
        3. Austrian: Red Bull Ring
        4. Azerbaijan: Baku
        5. Bahrain: Bahrain
        6. Bahrain: Sakhir
        7. Belgian: Spa
        8. Brazilian: São Paulo
        9. British: Silverstone
        10. Canada: Montreal
        11. Chinese: Shanghai
        12. French: Paul Ricard
        13. German: Hockenheimring
        14. German: Nürburging
        15. Hungarian: Hungaroring
        16. Italian: Monza
        17. Italian: Imola
        18. Italian: Tuscan (Mugello)
        19. Japanese: Suzuka
        20. Mexican: Mexico City
        21. Monaco: Monte Carlo
        22. Portuguese: Portimão
        23. Russian: Sochi
        24. Singapore: Marina Bay
        25. Spanish: Barcelona-Catalunya
        26. Turkish: Istanbul
        27. UnitedStates: Austin
        28. Others

Please enter the circuit:
Circuit
```

DSS UI, asking which racetrack to be predicted.

```
******************************
**Model 1: Pit Stop Strategy**
******************************

Predicting...

******************************
**Model 1: Pit Stop Strategy**
******************************

The given conditions are:

        Initial Position:  3.0
        Number of Pits:  2.0
        Tyre grid:  3.0
        Tyre 1:  2.0
        Tyre 2:  3.0
        tyre_grid_distance:  10.0
        tyre_1_distance:  25.0
        tyre_2_distance:  19.0
        Pit stops under safety car:  0.0
        Temperature:  75.0
        Humidity:  0.685
        Altitude:  10.7
        Turns:  21.0
        RaceDistance:  55.0
        TrackLength:  5.554
        TeamAbility:  131.0
        DriverAbility:  98.0

The predicted final position gained is: -0.0006594551 positions
Predicted final position is: 3rd

End of prediction, please enter the model to be used, or press ENTER to exit the program:

Model 1: Pit Stop Strategy
Model 2: Safety Car Scenario

Model
```

DSS UI, showing model 1 predicted results as 3rd place.



```
********************************
**Model 2: Safety Car Scenario**
********************************

Predicting...

********************************
**Model 2: Safety Car Scenario**
********************************

The given conditions are:

        Initial Position:  5.0
        Lap of safety car:  10.0
        Laps travelled:  10.0
        Laps remained:  61.0
        Current tyre compound:  2.0
        Current position:  2.0
        Temperature:  68.0
        Humidity:  0.68
        Altitude:  63.5
        Turns:  8.0
        RaceDistance:  71.0
        TrackLength:  4.318
        TeamAbility:  319.0
        DriverAbility:  214.0

If you choose not to go into the pit now, the predicted final position gained is: 2.0818138 positions
Predicted final position is: 3rd
If you choose not to go into the pit now, the predicted final position gained is: 1.3601308 positions
Predicted final position is: 4th
The model suggest that you should go into the pit now!

End of prediction, please enter the model to be used, or press ENTER to exit the program:

Model 1: Pit Stop Strategy
Model 2: Safety Car Scenario

Model
```

DSS UI, showing model 2 predicted suggestions to go into the pit.

The first model gives the user predictions over a given pitstop strategy to the race result, while the second model predicts under the accident scenario when the safety car is deployed, so that the team needs to react quickly whether to tell the racer to come into the pit to change a fresh new set of tyre compounds or stay out on the track. In our DSS design, a decision can be made within 15 seconds by typing in the required information, while in fact the prediction from the established machine learning model takes around 1-3 seconds only. Regarding improvement, after implementing the system to the team, the system can be merged into Scuderia Ferrari F1 Team's racing information system to automatically scrape the required input data from the live data database, such that a suggested pitstop strategy could be made within 4 seconds, which could greatly benefit race engineers and strategists.

Conclusion and Future Works

Conclusion

We have established our first model that predicts position change when given a pitstop strategy and track condition, under dry conditions. The first model gives us huge confidence that Scuderia Ferrari F1 Team could start relying on the prediction results and make more optimized pit stop strategies in future seasons. When testified by the 2020 season races, the model predicts the final position with ±2.2278 absolute position difference on average.

Our second model helps strategists make a quick decision during an Accident scenario with the safety car deployed. The team can swiftly react to the sudden potential pitstop decision by running the model for the predicted final position gain, in order to realize if the decision to pit in or stay out could give them the optimized race result. When testified by the 2020 season races, the model predicts the final position with ±2.0999 absolute position difference on average. This testing method would serve as the best examination method, since we cannot alter historical decisions and check the real-life result of "what if he did not pit in or did not stay out". However, we proved that by including more future race results in the model, the accuracy of the final position prediction would increase.

Our simplified Decision Support System functions fundamentally as we wished. By quickly gathering the required information from the user, the program promptly arrived at the pitstop strategy suggestions and their corresponding final position predictions. It could therefore effectively save time for the race strategists to form a final race strategy decision.

Improvement

However, the models still have areas to be improved. Firstly, without knowing the opponent's pitstop strategy decisions, the first model could only predict outcomes for a specific predefined scenario from racing strategists. In addition, wet weather conditions could imply different tyre strategies, which is not covered in our model training process. Regarding the optimal number of pitstops, from the 2016 season onwards, 1-stop and 2-stop pitstop decision predominates the races, and 3-stop strategies or more stops tend to give worse results. However, 3-stop might still be the optimal strategy especially under several safety car scenarios. Both of our models disregard successful or unsuccessful 3-stop pitstop strategies.

These mentioned scenarios are not taken into full consideration in the training process, hence unable to deliver predictions during wet conditions and more than 2-stop pitstop strategies, limiting the use cases of the system. More discussion on how to improve the model's accuracy would also be provided in the "future work" section below.

Future Works

Our future works include the followings:

Current Model improvement:

1. Improve Driver & Team Ability index score.
2. Confidence score for prediction results.

Model functionality extension:

3. Model for predicting opponent's pitstop strategies.
4. Model that predicts position gain or loss under wet conditions and DNF.
5. Data visualization & Dynamic Decision Support System (DSS).

Improve Driver & Team Ability index score

Currently, the model attempts to give credits to better constructor teams and drivers by an index score, so to differentiate competitive cars and racers from the weaker cars and inexperienced drivers. An intuitive way is to take the constructor championship scores and driver championship scores from the previous year. However, a driver's ability should be related to the experience the driver has and his driving style, and the fact that the driver's score from last season might be greatly affected by the car's competitiveness creates inaccuracy to the index score. An example would be Scuderia Ferrari F1 Team driver Charles Leclerc, who drove for Sauber Ferrari F1 Team (former of Alfa Romeo Racing), which has not been a competitive constructor team over recent years. To improve the driver ability index score, it is possible to take a driver's previous races in other championships, such as Formula 2 or Formula 3, as a reference to how the driver performs.

Throughout the years, constructor teams' competitiveness has also varied a lot. For Scuderia Ferrari F1 Team, it finished sixth place in the constructor championship in 2020, which is the worst season in 40 years: Even for the performance in 1980, it was a "sudden, brief dip" in the team's form, and were the championship the year before, and were back on top within two seasons (RaceFans, 2020). Our model suffers from prediction accuracy as a result of the sudden drop in Ferrari's car performance, hence we have to update the teams' ability index with the 2020 season constructor championship score instead. To solve this problem, we could take reference of the race pace of teams during each race qualifying session, where all cars will try to make the fastest lap time in order to start from front positions. In addition, free practice session could serve as a durability test, that could also contribute to the team ability index in the model.

Confidence score for prediction results

Our model could provide insightful predictions about final positions or pitstop strategy suggestions during the safety car scenario, with a model accuracy of approximately ±2 absolute position difference. However, it lacks a confidence score of how confident the prediction is. An improvement in providing the confidence level could help the well-informed race strategists form a better decision. To facilitate this feature, we may have to change the result label to a categorized result label, for example, a category for each integral final position. Next, we could use Explainable AI to run an analysis on the confidence level of each prediction for each category. The drawback of such an approach is that, when applied to model 2 about accident scenario, we cannot provide predictions on final position gain to 4 decimal points to compare whether staying out or pitting in would be a better strategy option. However, it still remains an immense reference value of how sure the model thinks about the prediction results.

Model for predicting opponents' pitstop strategies

Famous Chinese philosopher Sun Tzu once said, "know your enemy and know yourself, you will not be imperiled in a hundred battles." F1 pitstop strategy will not be complete if we fail to predict the opponents' strategies. However, due to the zero-sum game nature of Formula One, game theory suggests that different teams would react aggressively to opponents' racing strategies, including pitstop strategies, attacking or conservative driving, undercutting or overcutting, etc. This problem could not be simply solved by adopting a similar model like our two models, since inclusion of all opponent's racing strategies is required to understand how different racers interact with state changes and decisions made by opponents. A more sophisticated and advanced Artificial Intelligence model such as a multi-agent reinforcement learning model might be required for predicting how each racer as an individual agent should gain the optimal benefits by conducting what actions at what time.

Model that predicts position gain under wet conditions and DNF

Our model neglects all historical races under wet or mixed-weather conditions, which occasionally occur in a regular F1 calendar. While dry-weather conditions mainly challenge the car’s performance, durability, and the execution of pitstop strategies, wet-weather conditions testify drivers’ racecraft and skillsets, which include how well the driver handles cornering with much less grip between the tyres and the floor, and whether the driver can correct and save the car when it loses control.

Charles Leclerc losing grip and crashing to the barrier in the wet, Germany GP 2019

In addition to wet-weather conditions, if a racer crashed the car or retired from the race due to any reason, the label “DNF” which stands for did-not-finish would be applied to the record. Although in our model design we ignore such records and only focus on those with a race end result, the DNF records may stand for some implications, such as an over-aggressive tyre strategy that cause an overly degrading of tyres, that leads to spinning out off the track and crashing the car. However, there are also situations such as engine, suspension, or braking system failures, etc. These situations are uncontrollable regarding how the racing strategies are conducted, hence should not be included in the consideration. A throughout categorization of DNF records is needed to understand the cause of the car retirement, so that we could build a model that tries to avoid falling into the same situation of race retirement.

Data Visualization and Decision Support System (DSS)

Decision Support System is the interface for race engineers to evaluate and deploy the analysis results. The system should be simple, providing a set of neat input rows and a clear prediction result to update the team strategists effectively and efficiently. The user interface must be supporting real-time updates, highly interactive with high clarity. By updating the race condition in real-time, the system will simulate all possible pit-stop strategies at every lap dynamically and generate reports on each outcome. After ranking all outcomes, the system should suggest to the user on which lap should the car be pitted in for new tyres, with the probability of the predicted resulting track position.

By making a user interface with predictions and strategy suggestions, we could develop several easily interpretable data visualization tools such as interactive charts showing how different pitstop strategies affect different attributes such as final position, risk factors, or even to show comparisons of different suggested strategies.

From the literature review, we know the fatal mistake made in Abu Dhabi 2010 is due to the over-reliance on the system, and there are only 2 choices made by the DSS. In our design, we should emphasize providing a larger variety of optimal pit stop strategies and let the strategists more freedom to choose from several options.

References

Aversa, P., Cabantous, L., & Haefliger, S. (2018). When decision support systems fail: Insights for strategic information systems from Formula 1. The Journal Of Strategic Information Systems, 27(3), 221-236. doi: 10.1016/j.jsis.2018.03.002

Bekker, J., & Lotz, W. (2009). Planning Formula One race strategies using discrete-event simulation. Journal Of The Operational Research Society, 60(7), 952-961. doi: 10.1057/palgrave.jors.2602626

RaceFans. (2020). The remarkable stats of Ferrari's worst season for 40 years. Retrieved 17 April 2021, from https://www.racefans.net/2020/12/14/the-remarkable-stats-of-ferraris-worst-season-for-40-years/

Sulsters, C., & Bekker, R. (2018). Simulating formula one race strategies. Vrije Universiteit Amsterdam.